

# THIS WEEK

## EDITORIALS

**GREAT APES** Ethical concerns are key to discussions on chimp research **p.252**

**WORLD VIEW** Why the long arm of the law must still reach designer drugs **p.253**

**SLIP STREAM** Fishy information helps mullet motor along **p.254**



## Contaminated food for thought

*If it is to deal effectively with outbreaks of infectious diseases, Germany must streamline its convoluted systems for reporting and communication.*

Some six weeks after the first cases of potential food poisoning were reported, diners in Germany are still contemplating their side salads nervously, spooked by the confused information and warnings that have been issued over the past few weeks. Which item of greenery might be home to the deadly *Escherichia coli* bacterium known as EHEC O104:H4? By 13 June, the microbe had infected 3,325 people and killed 36.

The German public has been traumatized. It took weeks for the probable source of the bacterium to be named as an organic-bean-sprout farm in Lower Saxony. And, inevitably, accusations of crisis mismanagement are starting to fly.

These critical fingers, rightly, are not pointed at the scientists in Germany (and elsewhere), who rose admirably to the challenge of identifying and analysing the culprit. Instead, they are directed, with some justification, at the bizarrely complicated system Germany uses to handle disease outbreaks and track their sources — and at an alarmingly outdated way of transmitting information between physicians and agencies.

Ultimately responsible for disease control and prevention is the Robert Koch Institute in Berlin. However, Germany's federalized structure means that the institute receives its information indirectly, through many tiers of hierarchy.

The clinical laboratories that investigate samples sent to them by physicians and hospitals must promptly report notifiable diseases to their district health office, of which Germany has more than 400. Each of these offices passes the information on to its respective state ministry, which then transfers it to the federal health ministry, which then passes it onto its Robert Koch Institute. Days can elapse at transfer points and, scarcely credible in 2011, some of this information is still sent by post.

There is more. Responsibility to track the source of food-borne infections lies not with the Robert Koch Institute, but with the Federal Institute for Risk Assessment, part of the Ministry of Food, Agriculture and Consumer Protection. So, together there are two federal ministries, two federal technical institutes and 16 state ministries that can each pronounce on progress. Inevitably, confusion emerges — as demonstrated by the rushed and false fingering of Spanish cucumbers as the source late last month by Hamburg's state health minister, Cornelia Prüfer-Storcks.

Two things need to be done. First, Germany must eliminate the information-transfer chain and introduce a centralized electronic database that district health offices feed information into directly. Ideally, this would be supplemented by mandatory electronic reporting of individual cases by physicians. The US Centers for Disease Control and Prevention in Atlanta, Georgia, operates such a system, and the idea was discussed in Germany after the 2009 swine-flu pandemic. However, the proposal lost political support because it threatens the autonomy of the states.

This takes some explaining. Germany's post-war constitution was

designed to keep centralization to a minimum, and many responsibilities, including health, were devolved to the states. Introduced to prevent another dictator like Hitler, this principle is hard to attack. But it was never intended to hinder Germany from controlling politically illiterate microbes with no respect for state borders. Clearly, a way must be found to make an exception to the devolved-responsibility rule, at least when it comes to infectious diseases. The Robert Koch Institute, which has proven itself extremely competent in handling its part of the *E. coli* crisis given the blocks put in its way, needs much more power. Second, when disease threatens, Germany needs to be able to speak to its people with one voice — no matter how many authorities are involved in the process. This should be the Robert Koch Institute.

EHEC O104:H4 has proven to be a particularly evil enemy. Current agricultural practices are likely to generate other microbes of equal virulence or worse, and these will inevitably spread as people travel. Authorities in Germany and elsewhere must be able to keep control. ■

***"A way must be found to make an exception to the devolved-responsibility rule."***

## Full transparency

*Nations should release global nuclear-monitoring data to academics and the public.*

Under the auspices of a proposed international ban on all nuclear-weapons tests, scientists have built a system that can detect an illicit explosion anywhere in the world. The monitoring network stretches from Antarctica to Siberia and captures a wealth of useful data — not just on infrequent atomic bangs, but also on other types of explosion, earthquakes, underwater shocks and radiation releases.

Yet access to these data is restricted to contributing governments and selected allied scientists, who are largely prevented from sharing the information with the public. The diplomatic excuses offered for this unwise and unnecessary secrecy no longer wash, particularly in light of the March meltdowns at the Fukushima Daiichi nuclear power plant. At a meeting in Vienna next week, scientists who used these data to inform their governments about the scale and dangers of the Fukushima accident, but who saw the results kept under wraps, will push for change.

Their move deserves support. Data from the network, run by the Comprehensive Nuclear-Test-Ban Treaty Organization (CTBTO), should be freely available to scientists everywhere, for study in their own right and to inform the public in times of crisis. Governments

may be nervous about such openness, but the benefits far outweigh the risks.

The CTBTO has proved its worth in recent years. It detected North Korean nuclear tests in 2006 and 2009, and has captured detailed seismic data on major earthquakes, including the 2004 Sumatra–Andaman event that sparked a devastating tsunami.

This spring, the organization's 80 radioisotope-monitoring stations offered the clearest global picture of low-level fallout released from the Fukushima plant. Government-accredited scientific institutions were given access to provide politicians with valuable information about how the radiation was spreading and whether it posed a national threat. But most were told not to talk about the results in public, or to share the data with others in academia. The reason was diplomatic: governments such as the United States did not want to embarrass the Japanese, nor pre-empt their announcements about events unfolding at Fukushima Daiichi.

More generally, governments worry that radioisotope data are too sensitive to share. Politicians fear that, should a nuclear test occur, full access to incriminating data could somehow allow the offending nation to contest charges of weapons testing. Or perhaps that others could glean sensitive nuclear secrets from the isotopes in the atmosphere.

These fuzzy fears must be weighed against the impact of the information vacuum that followed Fukushima. Scientists everywhere were asked to give assessments, yet few had access to data that would allow them to do so. Providing open access to the CTBTO's network would

have given experts the information they needed to make important statements about Japan's reactors and the threats these posed to Tokyo and beyond. The data would also have lent credibility to the Japanese government's own statements on radiation levels in the region.

Moreover, such data are scientifically useful in their own right. Atmospheric scientists use radioisotopes widely and the CTBTO

***“Scientists were asked to give assessments of the fallout from the Fukushima plant, yet few had access to data that would allow them to do so.”***

network is gathering a unique data set that could be used to improve climate models or to refine meteorological studies. Scientists with access to the data might also find some new use for them. Thus far, nations have paid a combined US\$1 billion for the network, and they might as well put it to good use.

The network has already taken tentative steps towards openness. Following the 2004 tsunami, member states agreed that its seismic and hydroacoustic data could be used by accredited tsunami-warning centres around the world. In the immediate aftermath of the Fukushima Daiichi accident, it was allowed to share data with the International Atomic Energy Agency.

These are positive developments, but nations should go further: the CTBTO data are valuable in times of both calm and crisis. Contrary to the concerns of some, the more people who see them, the more valuable they will become. ■

## Great ape debate

*Researchers should contribute to a US analysis of the case for chimpanzee research.*

**T**he historical value of the chimpanzee as a disease model is indisputable. It was important in developing the Sabin polio vaccine; instrumental in discovering the infectious nature of the spongiform encephalopathies; and essential to both the creation of a vaccine against hepatitis B and the identification, in 1989, of the hepatitis C virus (HCV).

Humankind has benefited handsomely. Since the United States instituted universal childhood vaccination for hepatitis B in 1991, there has been a 98% decline in the disease in children under the age of 15 years. And with the identification of HCV, screening of donated blood for the virus reduced the risk of transfusion-associated hepatitis in the United States from 4% in 1989 to almost zero in 2000.

Today, chimpanzee research is still bearing fruit, especially for hepatitis C, a disease that infects at least 170 million people globally and often results in permanent liver damage or cancer. No approved vaccine yet exists. A study published in 2002 put the annual economic costs of the disease in the United States at more than US\$750 million.

The chimpanzee is the only animal model in which human strains of HCV can replicate, making it especially important in work to develop a vaccine. And studies in this animal have propelled at least one hepatitis C vaccine into human trials. Other chimpanzee experiments are making inroads in developing better therapies for the disease. The case for chimpanzee use in some other circumstances — such as the effort to develop a vaccine against respiratory syncytial virus, which mainly affects infants and young children — is less strong, but is at least arguable.

But chimpanzee studies are under fire (see page 268). Public discomfort over the use of chimpanzees in research has reached a historic high, with the result that the United States is now the only country save Gabon in which invasive experiments are conducted. Legislation has now been introduced in the US Congress that would prohibit invasive chimpanzee research. Although the bill is unlikely

to become law any time soon in a Congress distracted by wars, debt and a moribund economy, the Great Ape Protection and Cost Savings Act is nonetheless a sign of the times.

So, too, is the fact that the National Institutes of Health (NIH), facing public pressure after proposing to return nearly 200 semi-retired chimps to active research, has commissioned a study by an Institute of Medicine (IOM) committee, which convened last month. The committee's task, to culminate in a report planned for the end of the year, is to determine whether chimpanzee studies are necessary to answer current and future biomedical and behavioural research questions, or for drug and vaccine testing — and, if so, why.

The purview of the task that the NIH has set the IOM is troubling. It contains no mention of ethical aspects of the research, and the NIH has publicly stated that this omission was deliberate. Of the 12 current members of the committee, just one is a bioethicist. The agency may wish to divorce the science from the ethics, but society at large will not accept such a distinction. Nor is it intellectually defensible: a moral choice to use intelligent, emotionally complex creatures to their detriment, for the benefit of human welfare, is intimately related to what can be achieved scientifically. It would be wrong for the NIH to make any change in its support for chimpanzee research — or indeed to maintain the status quo — solely on the basis of the scientific report from the IOM.

Still, the work of the committee will provide a valuable starting point by defining the scientific case for chimpanzee research. Working from this, ethicists, the public, the animal-protection lobby, scientists and regulators could then engage in the much-needed, wider-ranging debate. An ideal convener for such a discussion would be the Presidential Commission for the Study of Bioethical Issues.

One thing is almost certain: if the NIH and scientists do not engage with the ethical and animal-welfare issues that are so clearly at the forefront of the public mind, Congress will do it for them, and the result may well be to shut down virtually all research using great apes, as happened in the European Union in 2010.

The committee plans to gather public input at a meeting in Washington DC in August, on a date yet to be announced. Researchers would do well to make their views known to the IOM committee, which will receive and consider all public comment at [go.nature.com/5tdgkt](http://go.nature.com/5tdgkt). ■

**➤ NATURE.COM**  
To comment online,  
click on Editorials at:  
[go.nature.com/xhunqy](http://go.nature.com/xhunqy)





## Poison in party pills is too much to swallow

*The harm caused by designer drugs justifies the law's attempts to keep pace with underground chemists, says Mike Cole.*

Europe's underground chemists have been busy. According to a report last month from the European Monitoring Centre for Drugs and Drug Addiction in Lisbon, the alarming rise in the appearance and abuse of drugs continues. In 2008, some 13 new substances were reported. In 2009, this figure rose to 24, and in 2010, the highest ever number of new drugs was reported, a total of 41 new substances.

Most were cathinones — related to amphetamines — or synthetic cannabinoids. All must have been made by skilled chemists as a deliberate challenge to drug-control laws.

Should we try to keep ahead of those who make and use these materials? Is the effort and expense required for chemists such as myself to develop tests for new drugs, and to work with legal professionals to increase the number of banned substances, really worth it? The simple answer is yes. I have seen the effects that these chemicals can have on those who take them. In addition to damaging medical conditions, these drugs can induce dangerous or violent changes in mood and behaviour. I believe that society has a duty to intervene.

Prosecuting a drug case requires a compound to be identified and shown to be illegal — not always an easy task. Some laws control drugs by name and chemical class, thereby outlawing a broad family of related chemicals. Others ban only specific isomers, opening the door to 'designer' drugs manufactured to mimic the effects but not the structure of an illegal compound. The Internet has facilitated international trade in these materials.

One such new drug is benzylpiperazine, or BZP. Developed by the Wellcome Research Laboratories in Beckenham, UK, in 1944 as an anthelmintic drug to combat parasitic worms in livestock, it was subsequently investigated as a potential antidepressant. It entered the party scene a decade or so ago as a legal alternative to ecstasy (MDMA), producing similar effects but not illegal at that time. Today, ecstasy tablets bought on the streets of London or San Francisco are as likely to contain BZP as MDMA.

BZP exemplifies the problems that new drugs and 'legal highs' pose for law-makers. Widespread use led to the compound being banned in the United States and much of Europe, yet it remains legal in other places, such as Canada.

Work in my laboratory has shown that BZP is not, as many users believe, safe. We treated immortalized cell lines from the liver and kidney — the excretory organs that clear drugs from the body — and fibroblasts, cells involved in wound-healing, with BZP, its precursor chemicals and its reaction by-products. We tested both the compounds and the impurities

created in their manufacture in isolation, as mixtures and as drug blends synthesized to mimic street samples. The concentrations used represented those recorded in the body during drug use.

All these chemicals were hugely toxic to both liver and kidney cell lines. The major impurity in BZP, dibenzylpiperazine, is especially toxic to the kidneys. One of the starting materials, piperazine hexahydrate, some of which can make it into the final product, is extremely toxic to the liver. These results start to explain the symptoms of renal and hepatic failure observed in people who use BZP.

Toxicity depends on the composition and concentration of the mixtures, and the effects are hard to predict. Other side effects include insomnia, anxiety attacks, nausea, vomiting and serious palpitations that frequently go unreported. These effects become worse when the

drugs are mixed with alcohol. In short, the effect on individuals is potentially significant, long-lasting and even fatal.

Control of such drugs brings its own problems. Synthesis of a compound is driven underground. BZP is easily manufactured from piperazine hexahydrate and benzyl chloride, but the level of impurities depends on the precise quantities of starting materials, the reaction conditions and the procedures used to extract the drug from the reaction mixture. This presents a paradox common in drug control: the safest option is for people not to ingest the chemicals, which is the aim of making them illegal. But making them illegal can make them more dangerous.

In response to this conundrum, people on both sides of the debate over whether to criminalize drugs often cite the economic benefit of their approach, but this argument is a red herring. Both sides have costs. Outpatient treatment after the ingestion of BZP costs hundreds of pounds per patient per visit. In-patient care, including treatment in an intensive-care unit, costs thousands of pounds a day. Society has a right to frown on and to seek to outlaw such costly behaviour. Yet the science behind a strategy of drug prohibition — quality-control methods in an analytical lab and access to forensic services — is expensive too.

So, returning to the original question: should we continue to outlaw recreational drugs, and compounds such as BZP in particular? The evidence is mounting that even pure drugs are toxic and do harm, both in the short and in the longer term. When public health and safety is at risk then surely it is socially responsible to ban these substances, and to provide a legislative and forensic-science system that supports such bans. ■

*Mike Cole is professor of forensic science at Anglia Ruskin University in Cambridge, UK.  
e-mail: Michael.cole@anglia.ac.uk*

**PROSECUTING A  
DRUG CASE REQUIRES  
A COMPOUND TO  
BE IDENTIFIED  
AND SHOWN TO  
BE ILLEGAL — NOT  
ALWAYS AN  
EASY TASK.**

➔ **NATURE.COM**  
Discuss this article  
online at:  
[go.nature.com/3svhpc](http://go.nature.com/3svhpc)

# RESEARCH HIGHLIGHTS

Selections from the  
scientific literature

## CANCER THERAPEUTICS

### Targeted drug fights melanoma

A drug that targets a specific mutant protein in skin cancer improved survival in a clinical trial of 675 patients with advanced melanoma.

The drug vemurafenib inhibits a mutated form of the cell-growth-promoting protein BRAF. Mutations in this protein are found in around half of all melanomas. Paul Chapman of the Memorial Sloan-Kettering Cancer Center in New York and his colleagues found that in their phase III trial of patients with metastatic melanoma and the BRAF mutation, almost half of those treated with vemurafenib responded to the drug. By contrast, the response rate in patients receiving an older chemotherapy called dacarbazine was only 5%.

Six months after treatment, 84% of those who received vemurafenib were still alive, compared with 64% of those who received dacarbazine. *N. Engl. J. Med.* doi:10.1056/NEJMoa1103782 (2011)

## PHOTONICS

### Rainbow from a single LED

Inorganic light-emitting diodes (LEDs) are bright, stable and efficient, but usually emit only one colour. Gyu-Chul Yi at Seoul National University and his team have created LEDs that can be tuned continuously from red to blue (**pictured**) for potential use in the display screens of mobile devices.

Their LED consists of

nanorods of the semiconductor gallium nitride, each coated with layers of indium gallium nitride. These layers form 'quantum wells' that restrict the movement of electrons, altering the electrons' energy levels and, ultimately, determining the wavelength of the LED's emitted light. The thickness of the layers varies naturally as they are deposited on the rods' multi-faceted tips. By altering an applied voltage, the researchers force electric current to travel through layers

of different thickness, thus changing the colour of light that the LED emits.

*Adv. Mater.* doi:10.1002/adma.201100806 (2011)

## NEUROGENETICS

### Extended hunt for autism genes

Boys are four times more likely than girls to have autism, and two studies hint at why: girls with the disorder tend to have many more genetic mutations

than boys, suggesting that girls undergo greater genomic change before showing autistic behaviour.

Groups led by Michael Wigler at Cold Spring Harbor Laboratory in New York and Matthew State at Yale University in New Haven, Connecticut, conducted the most comprehensive search yet for spontaneous duplications or deletions of stretches of DNA that may be associated with autism spectrum disorders. In analysing the genomes of more than 1,000 people — some with autism, some unaffected family members — the teams found at least 130 sites in the genome where spontaneous duplications or deletions might



A. REKKAS/ALAMY

## ANIMAL BEHAVIOUR

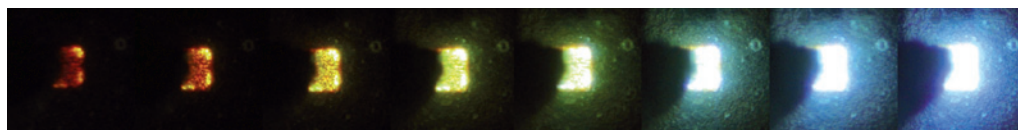
### Fitter fish lead the pack

Schooling fish take up different positions in the group according to their aerobic abilities.

Shaun Killen at the University of Glasgow, UK, and his colleagues noted the positions of individual juvenile mullet (*Liza aurata*; **pictured**) of similar size as the fish schooled in a swim tunnel in the lab, and measured certain animals' metabolic rates and swimming abilities. When schools were swimming at

high speed, fish less able to supply oxygen to their muscles ended up at the back, where they could reduce their workload. By contrast, fish with higher aerobic capacity that were better able to withstand drag forces took up positions at the front. Having fitter fish in the lead could allow schools to maximize their swimming speed.

*Proc. R. Soc. B* doi:10.1098/rspb.2011.1006 (2011)



WILEY-VCH

contribute to autism risk.

State's team found that duplication of a region on chromosome 7 is associated with autism. Autism is marked by antisocial behaviour, and deletion of the same region is linked to Williams–Beuren syndrome, a condition that involves hypersocial behaviour.

In a third study, Dennis Vitkup at Columbia University in New York and his colleagues, in collaboration with Wigler, analysed relationships between the mutated genes uncovered by Wigler's genetics study that were likely to be involved in brain function. Many clustered into a large network that regulates the creation and activity of connections between nerve cells.

**Neuron** 70, 863–885; 886–897; 898–907 (2011)

For a longer story on this research, see [go.nature.com/bscgfl](http://go.nature.com/bscgfl)

## GEOCHEMISTRY

## Mercury on the decline

A surprising drop in atmospheric mercury levels since the mid-1990s points to a substantial shift in the global biogeochemical cycle of the toxic element.

A team led by Franz Slemr of the Max Planck Institute for Chemistry in Mainz, Germany, compared data from monitoring stations in South Africa, Ireland and Antarctica, as well as measurements taken aboard ships in the Atlantic Ocean. They infer that, globally, mercury levels in the atmosphere have decreased by 20–38% since 1996.

Industrial mercury pollution has remained more or less constant over the past 15 years, leading the authors to suggest that decreasing re-emissions from soils and oceans of mercury deposited before the 1990s is the most likely cause of the downward trend. They add that climate change and ocean acidification may further shift the global mercury cycle.

**Atmos. Chem. Phys.** 11, 4779–4787 (2011)

## NEUROSCIENCE

## How nicotine curbs weight gain

Nicotine lessens the amount mice eat by activating specific neurons in the brain, perhaps explaining why people who stop smoking often gain weight.

Marina Picciotto at Yale University in New Haven, Connecticut, and her colleagues found that mice given nicotine daily for 30 days ate less and had lower body-fat levels than untreated mice. Nicotine increased the firing of brain neurons that produce a hormone precursor called pro-opiomelanocortin (POMC). When the POMC neurons fire, they release the hormone melanocortin. Mice in which the *Pomc* gene had been deleted ate the same amount whether or not they received nicotine. Those with a major melanocortin receptor gene silenced ate more when given nicotine than normal mice on nicotine.

The melanocortin hormone pathway regulates both energy use and food intake, so the authors think that nicotine has a two-pronged influence on body weight.

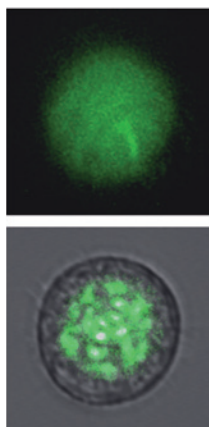
**Science** 332, 1330–1332 (2011)

## BIOPHYSICS

## Fluorescent cells turned into lasers

A human cell has been engineered to form the light source of a tiny laser — creating the first laser to use biological material to generate light.

Malte Gather and Seok-Hyun Yun at Harvard Medical School in Boston, Massachusetts, engineered human cells to express an enhanced version of green fluorescent protein. They then sandwiched a suspension of the cells between two tiny, closely



spaced mirrors to concentrate and align the light waves from the cells into a tight beam. By pulsing individual cells with blue light, the researchers excited the fluorescent proteins, causing them to emit light (two different lasing levels, **pictured**). The result was a bright directional beam of green laser light visible to the naked eye.

**Nature Photonics** doi:10.1038/

nphoton.2011.99 (2011)

For a longer story on this research, see [go.nature.com/iwdzj9](http://go.nature.com/iwdzj9)

## CHEMISTRY

## Recipe for a good catalyst

Faced with the challenge of developing low-cost catalysts for some fuel cells and metal–air batteries, researchers have come up with a basic recipe that ensures high catalytic activity in a family of widely used materials.

Perovskite oxides catalyse

the oxygen-reduction reaction, a core process in fuel cells and batteries. Yang Shao-Horn and Hubert Gasteiger at the Massachusetts Institute of Technology in Cambridge and their group studied 15 different perovskite oxide materials, which contain transition-metal ions. They found that the materials' catalytic activity in reducing molecular oxygen is strongly dependent on the level of occupancy of the transition metal's  $e_g$  electron orbital.

Because of various electron interactions between atoms of the oxide, this occupancy level can vary between 0 and 2. With one electron in this orbital, catalytic activity increased by four orders of magnitude compared with oxides that have 0 or 2 electrons in the orbital. An occupancy of less than 1 led to an interaction with the incoming oxygen that was too strong, whereas occupancy of greater than 1 made it difficult for the catalyst to interact with and adsorb the molecule.

**Nature Chem.** doi:10.1038/nchem.1069 (2011)

➔ **NATURE.COM**

For the latest research published by Nature visit:

[www.nature.com/latestresearch](http://www.nature.com/latestresearch)

## COMMUNITY CHOICE

The most viewed papers in science

## METAGENOMICS

## Movies of the body's bacteria

**HIGHLY READ**  
on genome  
biology.com  
7 May–6 June

The rich microbial populations of the human body shift frequently in time — even on a daily basis — with no stable, core group of microbes present at high levels.

Rob Knight at the University of Colorado in Boulder and his colleagues obtained daily microbial samples from the faeces, mouth and palms of two volunteers, over 6 months for one subject and 15 months for the other. The authors sequenced a key genomic region of the bacteria to assess the composition of taxonomic groups. This revealed that the microbial communities are distinct from one body site to the next both in each individual and between individuals. However, only a small proportion of the observed groups persisted across all time points. The authors suggest that factors such as diet, medication and differences in immune-system activity may explain the temporal variations.

**Genome Biol.** 12, R50 (2011)

For a longer story on this research, see [go.nature.com/iwdzj9](http://go.nature.com/iwdzj9)

For the latest research published by Nature visit: [www.nature.com/latestresearch](http://www.nature.com/latestresearch)



# SEVEN DAYS

The news in brief

## POLICY

### No nuclear in Italy

The Italian government has had to kill plans to reintroduce nuclear power, after a referendum on 12 and 13 June voted overwhelmingly to keep the nation nuclear-free. All nuclear power stations in the country have been closed since 1990, as a result of a referendum held after the 1986 Chernobyl disaster. Italy is already noted for its high dependence on imported electricity, which amounted to 14% of electricity demand in 2009, according to Italian utility company Terna.

### Arctic agreement

A barrier to exploration for oil and gas in Arctic waters was removed on 7 June when Norway and Russia ratified a deal on how to share the Barents Sea region, which is potentially rich in fossil fuel. In September 2010, after a four-decade dispute about the dividing line, the two nations' foreign ministers signed a treaty in Murmansk that split the Barents Sea equally. The treaty will be implemented from 7 July.

## RESEARCH

### Bean sprouts: guilty

The source of the *Escherichia coli* outbreak that has swept across Europe over the past month has been identified: bean sprouts. The Robert Koch Institute, the German federal agency for disease surveillance in Berlin, confirmed the culprit on 10 June. By 13 June, 36 people around the world had died and 3,324 had been infected. See Editorial, page 251.

### Tevatron clash

Two research groups at the Tevatron, the particle collider at Fermilab in Batavia, Illinois,



WHO

## Vaccine provides hope for meningitis

A cheap vaccine that was rolled out in three African countries in December has scored an early success. Burkina Faso, Mali and Niger have all reported the lowest number of meningitis A cases ever recorded in an epidemic season, six months after 20 million people received the MenAfriVac vaccine (pictured; see *Nature* 468, 143; 2010). No one immunized

is known to have contracted the bacterial disease, which periodically kills thousands during intense epidemics. The Meningitis Vaccine Project — led by the World Health Organization and PATH, a non-profit body based in Seattle, Washington — plans further immunizations, starting with Cameroon, Chad and Nigeria later this year.

disagree about whether they have spotted new particles. In the past couple of months, researchers on the Collider Detector at Fermilab experiment reported evidence of particles not predicted under the standard model of particle physics. But on 10 June, researchers on the independent D0 experiment said that their data do not confirm the signal. The two teams rarely disagree. See [go.nature.com/t46nns](http://go.nature.com/t46nns) for more.

### China's Moon probe

China's lunar orbiter Change-2 has left the Moon and is heading out into the Solar System, state news media reported on 9 June. The unmanned probe had been taking high-resolution images

of the Moon's surface as part of China's preparation for a future lunar rover. Change-2 is now headed for L2, a point beyond the Moon's orbit where the gravitational pull of the Sun and Earth are equal. Its arrival there in September will mark the farthest that China has ever sent a satellite.

### Frog fungus

Chytridiomycosis, a virulent fungal disease of amphibians, now affects the entire mountainous neotropics of Central America. On 13 June, scientists at the Smithsonian Tropical Research Institute in Panama announced that they had found infected frogs at a site bordering the Darien National Park, previously the only area to

escape the infection. The area is considered the "best shot" for biologists hoping to collect frog species to preserve in captivity, said Brian Gratwicke, a biologist at the Smithsonian Conservation Biology Institute in Washington DC. "We would like to save all of the species in the Darien, but there isn't time to do that now," he said. See [go.nature.com/pve1su](http://go.nature.com/pve1su) for more.

### Age of Aquarius

After two failed launches for NASA's Earth observation projects, the space agency has a success: Aquarius, its satellite to measure the saltiness of the oceans, reached orbit on 10 June. The probe will pick up weak microwave radiation emitted naturally by the

SOURCE: FAO

ocean. This radiation varies according to the electrical conductivity of the water, which in turn is tied to its salinity. Because salinity is linked to evaporation and water density, the data could help scientists to confirm theories about the global water cycle and its response to climate change. See [go.nature.com/jfwfhf](http://go.nature.com/jfwfhf) for more.

## Bias in science

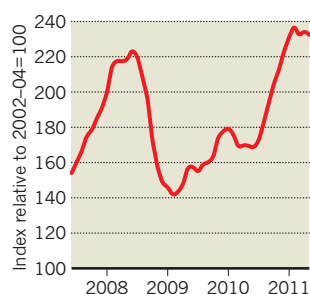
The eminent evolutionary biologist and science historian Stephen Jay Gould may have fudged his numbers, when criticizing skull measurements by nineteenth-century American physician Samuel Morton as a classic example of bias influencing scientific results. Gould — who died in 2002 — made the charges in 1978 (S. J. Gould *Science* **200**, 503–509; 1978). But, in a 7 June paper, a group of anthropologists argues that most of Gould's criticisms are “poorly supported or falsified” (J. E. Lewis *et al.* *PLoS Biol.* **9**, e1001071; 2011). See [go.nature.com/r1szy4](http://go.nature.com/r1szy4) for more.

## BUSINESS

## High food prices

World food prices are likely to remain “high and volatile”, according to the Food and Agriculture Organization. The agency warned in a 7 June update that its food

## FOOD PRICE INDEX



price index for May was only slightly down from a record level in February (see **chart**). Unfavourable weather, rising oil prices, political unrest and the nuclear disaster in Japan have all helped to unsettle the food market. The higher prices have boosted planting, but this year's harvests — particularly for crops with depleted stocks, such as maize (corn) — will be crucial, the agency said.

## Chinese windfalls

Huaneng Renewables, the wind-energy subsidiary of China's largest power producer, started trading last week after raising HK\$6.23 billion (US\$800 million) in an initial public offering (IPO) in Hong Kong. The firm had shelved its IPO in December because of a lack of interest. Its re-entry into the market follows January's 9.5 billion renminbi (US\$1.4 billion) IPO of Sinovel, China's largest, and the world's second-largest, maker of wind turbines.

## FUNDING

## Physics in Jordan

A US\$110-million synchrotron seems to be on track for construction in Amman, Jordan — surviving a global recession, political upheaval and the assassinations of two members of the project's Iranian delegation. The Synchrotron-light for Experimental Science and Applications in the Middle East (SESAME) project has received commitments from Israel, Iran, Jordan and the Palestinian Authority to provide funding as long as two further nations commit funds — as Egypt and Turkey are expected to do. Chris Llewellyn Smith, a physicist at the University of Oxford, UK, and president of the SESAME council, says that he is “confident” that the project will deliver its first three beamlines by 2015.

## California cash

The W. M. Keck Foundation in Los Angeles, California, has announced the single largest science donation in its 57-year history: a US\$150-million gift to the University of Southern California, Los Angeles, for scientific research at its medical school and two affiliated hospitals. The foundation also gave \$110 million to the university's medical school in 1999. For

## COMING UP

### 17 JUNE

The European Commission reveals the results of a vote on the new name of Europe's post-2013 research funding system.

[go.nature.com/oxsurs](http://go.nature.com/oxsurs)

### 20–25 JUNE

In Lima, Peru, scientists preparing the Intergovernmental Panel on Climate Change's fifth report meet to discuss geoengineering, and the ethics and economics of estimating the ‘cost’ of climate change.

[go.nature.com/jfc8le](http://go.nature.com/jfc8le)

### 20–24 JUNE

The International Atomic Energy Agency holds a meeting on nuclear safety in Vienna, Austria, to identify lessons from the Fukushima disaster.

[go.nature.com/jwjewx](http://go.nature.com/jwjewx)

the university, the donation on 13 June follows shortly after a record gift of \$200 million for science research, in March (see *Nature* **471**, 271; 2011).

## Vaccine cash bounty

The GAVI Alliance, a global health partnership that focuses on getting vaccines into low-income countries, has been promised US\$600 million more funding than it expected at a meeting in London on 13 June. The alliance, based in Geneva, Switzerland, needed \$3.7 billion to enable a planned \$6.8-billion expansion of vaccination programmes in 2011–15, yet donors committed \$4.3 billion. The pledges include an extra \$1.3 billion from Britain and \$1 billion from the Bill & Melinda Gates Foundation. See [go.nature.com/qlldf4](http://go.nature.com/qlldf4) for more.

**NATURE.COM**

For daily news updates see:

[www.nature.com/news](http://www.nature.com/news)

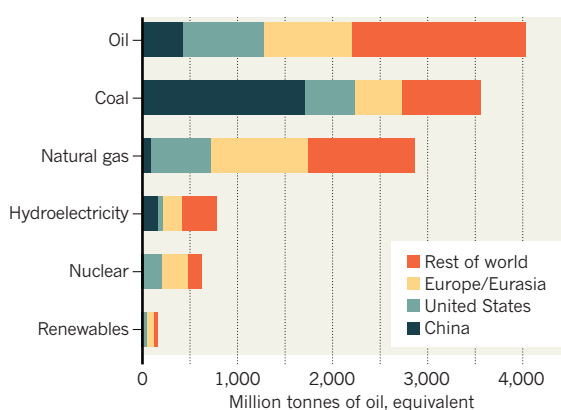
SOURCE: BP

## TREND WATCH

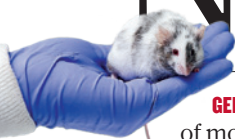
China accounted for just over one-fifth of the world's energy consumption in 2010, and 48.2% of global coal consumption, according to statistics released by oil company BP on 8 June (see **chart**). Oil is still the world's leading fuel (33.6%), although its market share has dropped about 5% over the past decade, whereas renewables make up just 1.3%. Overall, however, every fuel is being used more than ever before, as energy use has rebounded from a recession-induced drop in 2009.

## WORLD ENERGY USE 2010

World energy consumption rose by 5.6% from 2009; fossil fuels' contribution remained almost static, at 87% of the total.



# NEWS IN FOCUS



**GENOMICS** Library of mouse knockouts nears completion **p.262**

**EUROPE** Millions of euros siphoned off for fake research projects **p.264**

**EGYPT** Rising budgets and a warmer climate buoy scientists **p.266**

**QUANTUM MECHANICS** Spooky physics finds its way into biology **p.272**

WELLCOME LIBRARY, LONDON

AP PHOTO/J. ROBERSON



Devastating floods on the Mississippi this year have given researchers a rare opportunity to study how the river deposits the muds that build coastal marshes.

## SEDIMENTOLOGY

# Studies spy on a river's rage

*Investigation into this year's Mississippi floods could shape coastal restoration plans.*

BY GWYNETH DICKEY ZAKAIB  
IN VENICE, LOUISIANA

The research vessel *Acadiana* rolls with the waves in the Gulf of Mexico, 10 kilometres off the coast of Louisiana. Scientists and crew members scan the murky waters. Suddenly, triggered by an acoustic signal, a cluster of bright-yellow buoys comes bobbing up to the surface.

The captain steers towards the floats, which carry a radar instrument that has spent the past 20 hours on the sea floor. The device has been measuring the velocity of the water pouring from the Mississippi River, where floodwaters have risen to levels not seen in decades. "There's

a raging torrent coming out," says Carol Lutken, associate director for research programmes at the Mississippi Mineral Resources Institute in Oxford, Mississippi, which helped to organize the expedition. "It's like a fire hose."

The survey earlier this month is part of an ongoing interdisciplinary effort by researchers to learn how the flooding river discharges water and where it deposits its sediment load. Those muds could have a role in restoring the diminishing marshes along the Louisiana coast. The flood "is a catastrophic event, but it's a rare opportunity to understand the physics of the Mississippi delta," says Federico Falcini, a physical oceanographer at the University of Pennsylvania in Philadelphia, and

one of the project coordinators.

Heavy rains across the Mississippi watershed in April led to devastating floods far up the waterway, forcing entire communities to evacuate. On 14 May, as the high water moved into Louisiana, the US Army Corps of Engineers began opening floodgates in the Morganza Spillway, some 450 kilometres upstream of the gulf. Their purpose was to divert water into the Atchafalaya River, which follows its own course to the sea. The move spared developed areas downstream — including the cities of Baton Rouge and New Orleans. It also set up the ideal conditions for a direct comparison of river dynamics and sediment deposition in two very different waterways. ▶



► The Mississippi is hemmed in by a system of embankments and kept open for shipping by constant dredging. It reaches the sea in a few long, narrow channels, flanked by diminishing marshlands. By contrast, the less-controlled Atchafalaya emits a diffuse plume, which exits through networks of bifurcating channels and feeds a growing marsh called the Wax Lake Delta. Using satellites to compare the discharge of the two rivers, and the *Acadiana* to verify the satellite measurements of the Mississippi, Falcini and his colleagues will try to determine the conditions that build healthy wetlands. “The theory is, if you can tune the channel geometry on the Mississippi River Delta, maybe it will do something like what the Wax Lake is doing: spreading and making deposition just in front of the river,” says Falcini.

He and Douglas Jerolmack, a geophysicist at the University of Pennsylvania, hope to validate a model in which the faster water is moving as it exits a river, the farther into the sea it will carry sediment (F. Falcini and D. J. Jerolmack *J. Geophys. Res.* doi:10.1029/2010JF001802; 2010). This relationship would be especially important during floods, which carry unusually heavy loads of sediment that contribute to marsh-building along the coast. “We’re going to learn a whole lot that we’ll use to inform and expand our model,” says Jerolmack.

Falcini’s team is one of a few groups studying

the floods for clues to restoring the wetlands, which buffer the coast from hurricanes, floods and storm surges by absorbing water and wave power. The US Geological Survey says that nearly 43 square kilometres of Louisiana’s coastal marshes have been disappearing each year since 1985, owing to sea-level rise, subsidence and sediment deficit.

As a result, the open waters of the gulf are creeping closer to New Orleans, increasing its vulnerability to hurricanes. “We’re conscious that if we don’t do something to save the landscape, the entire coast will be gone,” says Steve Mathies, executive director of the Louisiana Office of Coastal Protection and Restoration in Baton Rouge, which is developing a plan for coastal restoration.

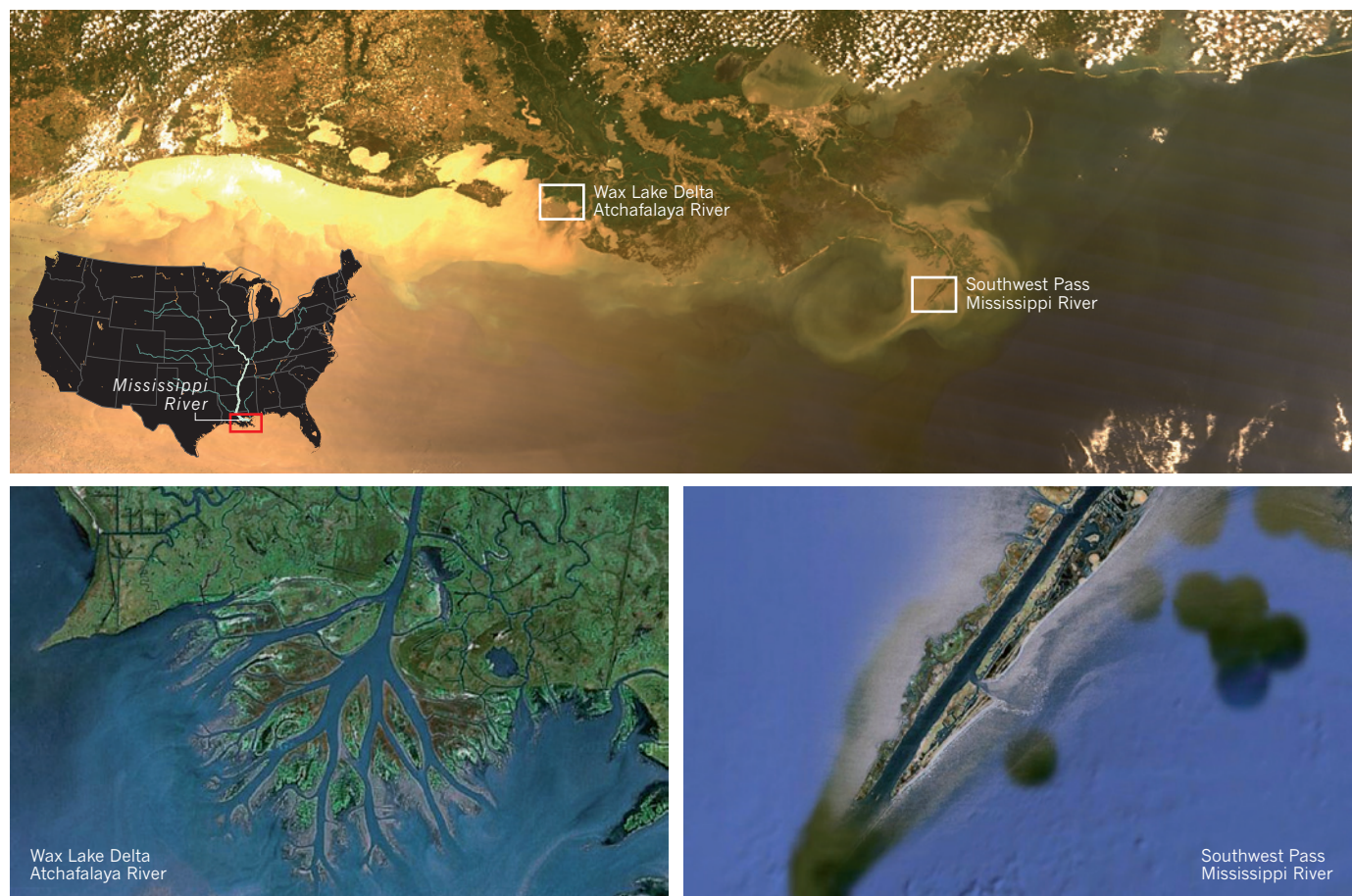
Some scientists say that the best way to save the coast is to divert more of the Mississippi’s floodwaters upstream to increase the amount of sediment reaching the marshes, but they don’t know how to ensure that the sediment will be captured where it is needed. Falcini’s model suggests that one way would be to widen channels that feed into the ocean, so that the water slows down and sediments settle out.

The researchers are using satellites to track sediment concentration at the water’s surface,

among other variables. They will validate those data using velocity measurements and samples gathered by the *Acadiana*. In the coming weeks, the floodwaters will subside, and project collaborators will fly a helicopter along the 250 kilometres or so of coastline between the Atchafalaya and Mississippi deltas, landing at regular intervals to take sediment samples and discover where the flood has added or eroded soil. Taken together, these measurements should help to determine which river characteristics can be tuned to build healthy marshes. Independent teams working farther up the river will add detail.

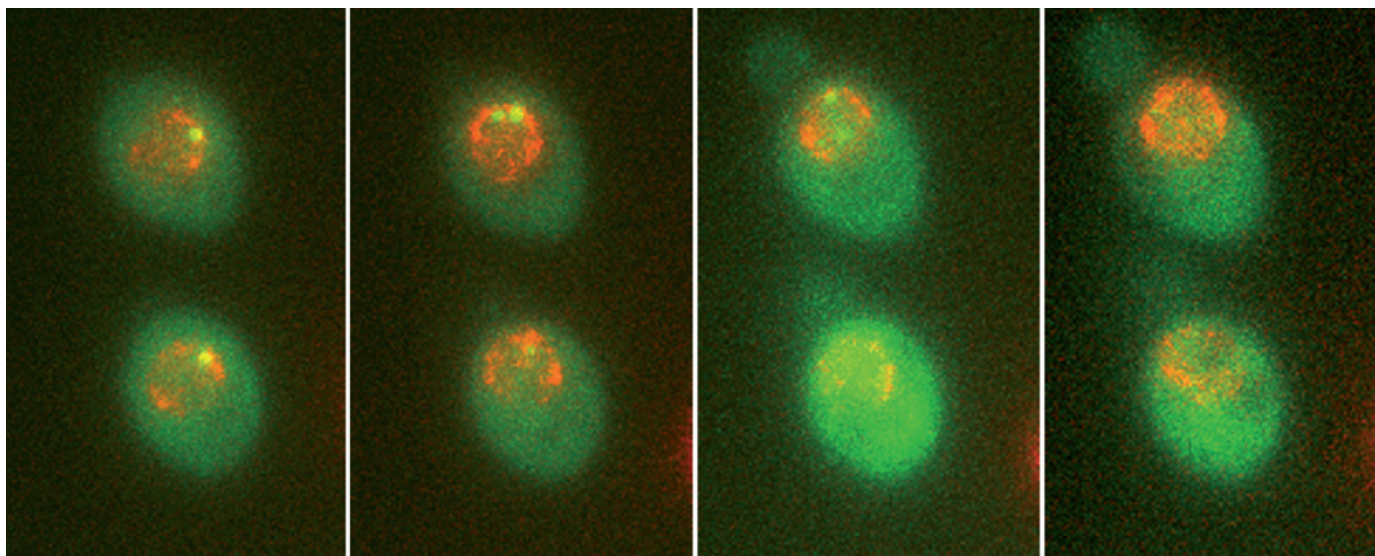
Besides shaping restoration plans, says Mathies, projects such as Falcini’s could also help to build support for restoration measures, which will be expensive and could meet resistance from stakeholders such as fishermen and shipping interests. “We need to be able to show people what the returns will be,” says Mathies.

Louisiana officials have started to take a “generational view” of coastal restoration, accepting that long-term benefits trump short-term interests, says Robert Twilley, a coastal-systems ecologist at the University of Louisiana at Lafayette. The trick to maintaining a healthy coastline is to minimize the damage from floods while maximizing the benefits, he adds. “We need to think how we can reconfigure the river to accomplish both flood control and restoration.” ■



Top: Sediment pours from the Atchafalaya on 1 June in a more diffuse flow than from the faster Mississippi. Bottom: The rivers have strikingly different deltas.





Real-time observation of yeast genes tagged to fluoresce when transcribed into RNA could help synthetic biologists to design better circuits.

## SYNTHETIC BIOLOGY

# Life hackers seek new tools

*Field aims to enlist techniques from molecular biology to attack fundamental challenges.*

BY ERIKA CHECK HAYDEN

Julius Lucks has heard the criticisms of synthetic biology before: life is too complicated to be manipulated by human designers; those who try have managed to cobble together only rudimentary genetic circuits from a limited suite of parts; the results are notoriously unpredictable. Meanwhile, a few high-profile successes — such as last year's creation of a bacterium with a synthetic genome<sup>1</sup> — and enthusiastic claims that the field will solve a raft of complex health, environmental and engineering problems, only increase the pressure to deliver.

Lucks, however, is undaunted. Last month he, his wife and their young child moved across the United States, from Berkeley, California, to Ithaca, New York, where he will set up his first independent lab in the discipline at Cornell University. His optimism is representative of a new generation of synthetic biologists who are gathering to chart the course of their field this week at a conference at Stanford University in California.

Jeff Tabor, a bioengineer at Rice University in Houston, Texas, says that one goal of the conference, the fifth Synthetic Biology Meeting, is to bring more traditional molecular biologists “into the fold”, both to counter their intrinsic

resistance to the concept of re-engineering life and to co-opt their tools. “There is a real difference in the way that I and people younger than me

see biology and think about studying cells,” Tabor says, “but there are a tonne of scientists doing molecular biology work that is improving our ability to engineer biology.”

For example, ‘next-generation’ sequencing machines, designed to vastly speed up the reading of genomes, can also offer synthetic biologists a better way to observe cellular behaviour. That in turn will help them design better circuits — for instance, by giving them a quantifiable readout of how a circuit's modifications affect its function. In a paper<sup>2</sup> published this month by Lucks

and his colleagues at the University of California, Berkeley, the group inferred the three-dimensional shapes of small RNA molecules by sequencing the corresponding DNA, using a technique called SHAPE-Seq. That strategy could help synthetic biologists to screen large pools of RNA rapidly to find those with certain structural characteristics that could be incorporated into RNA circuits.

Another tool that synthetic biologists hope to adopt was published in April<sup>3</sup> by a team led by structural biologist Robert Singer of the Albert Einstein College of Medicine in New York. By tagging particular genes with a signalling molecule that fluoresces every time the gene is transcribed, the researchers can watch and quantify transcription in real

time. The work could give synthetic biologists the equivalent of an electrician's circuit tester, helping them to engineer more predictable biological circuits.

“Using a technology like this, you can see exactly what a circuit is doing and count the number of circuit signals that are being produced in real time in live cells,” Tabor says. “This is exactly what we need to help us put circuits together.”

Bioengineer Adam Arkin, Lucks' mentor at Berkeley, has pursued the idea that circuits can be made more reliable by basing parts on existing cellular components that already accomplish a certain function in the cell. Such ‘mother parts’ could be tweaked slightly to yield ‘families’ of parts with similar features that could carry out their functions independently and efficiently.

In April, the team published a proof of concept for this approach<sup>4</sup> in which they tweaked an RNA-based gene-regulation system to simultaneously control the expression of multiple genes in a cell from the bacterium *Escherichia coli*, and even make a simple RNA circuit. Because the system is entirely RNA-based, it eliminates the need to translate a messenger RNA into a protein regulator, thereby reducing the overall complexity of the system.

Another approach to complexity involves designing multicellular circuits in which each cell is a circuit component. This neatly skirts the dilemma of trying to insulate the parts of a circuit from one another within the cytoplasm of a single cell. Chris Voigt, ►

**“There needs to be a frank and open discussion about funding in synthetic biology.”**

► NATURE.COM

For more on synthetic systems biology, see: [go.nature.com/dq38zq](http://go.nature.com/dq38zq)

► who is moving from the University of California, San Francisco, to co-direct a new synthetic-biology institute at the Massachusetts Institute of Technology in Cambridge, has been pursuing this approach in his lab with colleagues who published their proof of principle last December<sup>5</sup>. “There’s been a change in the scale of the problems that we can address, and this comes out of the tools that synthetic biology can provide,” says Voigt. At the meeting, Voigt will describe his lab’s attempts to re-engineer the way some organisms convert nitrogen into a useable form through a molecular pathway that involves dozens of genes.

Synthetic biologists have been organizing their own initiatives to tackle other obstacles. For instance, one of the field’s key tenets is that off-the-shelf molecular ‘parts’ could be used to program cells to carry out specific functions, such as making a drug or a biofuel. But such ambitious goals depend on the quality of the available parts<sup>6</sup>. So, in late 2009, an initiative called the BIOFAB (International Open Facility Advancing Biotechnology), funded by the US National Science Foundation, began working to design reliable parts with known functions. The BIOFAB has now made about 3,000 well-characterized parts and has released around 500 as a higher-quality curated collection.

Yet money is scarce for this kind of work — a challenge to be addressed at a conference workshop that will include funding agencies and industry. “There needs to be a frank and open discussion about funding in synthetic biology, especially in the United States,” says Pam Silver, a systems biologist at Harvard University in Boston, Massachusetts. The bread-and-butter work that the field needs, such as fine-tuning circuitry, is more applied than most ‘hypothesis-driven’ research that is the remit of agencies such as the US National Institutes of Health. And most funders want applicants to focus on specific agendas, such as health or biofuels.

Indeed, Rob Carlson, a principal at the engineering, consulting and design company Biodesic in Seattle, Washington, wonders whether the field of synthetic biology is big enough to become a well-oiled engineering machine. This week’s conference is sold out at 700 attendees, with a waiting list of at least 100, but as Carlson points out, many of those attending will be reporters and investors.

“Given the complexity of the task at hand, it doesn’t surprise me at all that we are still going slowly,” says Carlson. ■

1. Gibson, D. G. *et al. Science* **329**, 52–56 (2010).
2. Lucks, J. B. *et al. Proc. Natl Acad. Sci. USA* doi:10.1073/pnas.1106501108 (2011).
3. Larson, D. R., Zenklusen, D., Wu, B., Chao, J. A. & Singer, R. H. *Science* **332**, 475–478 (2011).
4. Lucks, J. B., Qi, L., Mutalik, V. K., Wang, D. & Arkin, A. P. *Proc. Natl Acad. Sci. USA* **108**, 8617–8622 (2011).
5. Regot, S. *et al. Nature* **469**, 207–221 (2011).
6. Kwok, R. *Nature* **463**, 288–290 (2010).



Mutant mice generated from embryonic stem-cell lines should further understanding of human disease.

#### GENOMICS

# Mouse library set to be knockout

*Global effort to disable every mouse gene nears completion.*

BY ELIE DOLGIN

Investigators are on the home stretch of the largest international biological research initiative since the Human Genome Project. Launched in 2006 in North America and Europe, the effort aims to disable each of the 20,000-odd genes in the mouse genome and make the resulting cell lines available to the scientific community.

After five years and more than US\$100 million, the pace is picking up. “In the next three years or so we assume we will have it completed,” says Wolfgang Wurst, director of the Institute of Developmental Genetics at the Helmholtz Centre Munich in Germany and one of the leaders of the effort’s European contribution.

“This resource will be of enormous benefit, not just to the mouse genetic community but to every scientist, every company looking at mammalian physiology, and of course everyone who wants to design better drugs and better health care,” says Steve Brown, director of the Mammalian Genetics Unit at MRC Harwell, UK. “It is one of the most significant biological resources in

the past century of science, and I don’t think I’m overstating the case here.”

Previously, researchers typically spent years engineering mice to lack specific genes so that they could model human diseases involving those genes. This process was slow, laborious and piecemeal. And even after all that effort, there was often no easy way to share the animals with other researchers. So the International Knockout Mouse Consortium (IKMC) set out to create a library of mouse embryonic stem-cell lines representing every possible gene knockout, and then to distribute the cells to researchers for further study.

A new technology — pioneered by Bill Skarnes and Allan Bradley at the Wellcome Trust Sanger Institute in Hinxton, UK, and described today in *Nature* (W. C. Skarnes *et al. Nature* **474**, 337–342; 2011) — helped make that possible. Using a high-throughput gene-targeting pipeline that allowed them to precisely engineer hundreds of genes every month, the Sanger team, in collaboration with colleagues in Germany and the United States, has so far inactivated more than 9,000 genes in mouse embryonic stem cells. It is on track to knock out 7,500 more in the next few years. “We’re really hitting our peak production now,” Skarnes says.

► NATURE.COM

For more on the mouse genome, see: [go.nature.com/4iifq1](http://go.nature.com/4iifq1)



Each bespoke knockout in the Sanger group's library contains an added 'conditional allele'. This allows scientists to disrupt gene function in a living mouse at any body site and at any point in the animal's development by the timely addition of enzymes that recognize the inserted allele. By this means, the effects of the missing gene do not kill the mouse before the researchers have a chance to study it.

"It is truly a feat of genius," says Geoff Hicks, a geneticist at the University of Manitoba in Winnipeg who leads the Canadian contribution to the IKMC. "This paper really pushed the technology in an extremely innovative way and met a challenge that seemed unattainable."

Various groups in the international effort are using other, non-conditional techniques to inactivate thousands more genes. Researchers in Texas, Canada and Germany have mutated close to 12,000 genes using an untargeted approach called gene trapping, and Regeneron Pharmaceuticals, a company based in Tarrytown, New York, has specifically targeted around 3,500 genes using a technology that works well in smaller genes but results in mice that are less flexible for research than conditional knockouts. "The approaches are complementary," says Aris Economides, Regeneron's senior director of genome engineering technologies. "This is going to play out well for the end user."

To date, nearly 17,000 different genes have been knocked out, leaving only around 3,000 more to go. The Sanger team, however, hopes to replace most of the genes hit by gene trapping with conditionally targeted knockouts, because targeting allows individual genes to be manipulated with greater precision.

Already, mutant mice have been generated from almost 1,000 of the embryonic stem-cell lines obtained, and the IKMC repositories in the United States, Canada and Europe receive hundreds of new orders every month. The next challenge is to study the function of each missing gene. To this end, the US National Institutes of Health last year committed \$110 million over the next five years to characterize around 2,500 of the IKMC's mutant mice through the International Mouse Phenotyping Consortium, with plans for another \$110 million to define 5,000 more if the first phase is successful.

"Knocking out the mice is simple relative to the huge task of finding out what all those genes do," says Richard Finnell, a geneticist at the Texas A&M Health Science Center in Houston. ■

**"It is one of the most significant biological resources in the past century of science."**

## BOTANY

# Species spellchecker fixes plant glitches

*Online tool should weed out misspellings and duplications.*

BY JOHN WHITFIELD

Brian Enquist and his collaborators were delighted with their freshly compiled data set of 22.5 million records on the distribution and traits of plants in the Americas. But their delight turned to horror when they realized that the data set contained 611,728 names: nearly twice as many as there are thought to be plant species on Earth.

Completed in December 2010, the records were intended to help Enquist and his colleagues to discern trends in how forest trees in a wide variety of environments respond to climate change. But the data were clearly full of bogus names, making it impossible to count the species in a particular area, or their relative abundance. "I started to question our ability even to compare something as basic as species diversity at two sites," says Enquist, a plant ecologist at the University of Arizona in Tucson.

This month, Enquist's team will unveil a solution that could help botanists and ecologists worldwide. The Taxonomic Names Resolution Service (TNRS) aims to find and fix the incorrect plant names that plague scientists' records.

"It looks really good," says Gabriela Lopez-Gonzalez, a plant ecologist at the University of Leeds, UK, who curates a database of forest plots. Fixing species lists by hand is arduous, she says. "This should save us a lot of time".

She and others agree that the problem is widespread in botanical databases. "Digitization has made the problem worse," says TNRS co-leader, botanist Brad Boyle, also at the University of Arizona. Boyle explains that as more data are added to digital records, the chance of introducing errors also increases. Even in herbarium specimens, which ought to be the gold standard for plant identification, about 15% of the names are misspelt, he says.

Many of the errors seem to arise because biologists are not as careful as they should be when entering data into digital records. The TNRS team estimates that about one-third of the names entered into online repositories — such as GenBank, the US National Institutes of Health collection of DNA-sequence data, or the Ecological Society of America's VegBank database of plant-plot data — are incorrect.

The other problem is that names change. Old names can be abolished when experts reclassify plants as ideas about evolutionary relationships change, or when they realize the species already



R. EVANS/PHOTOLIBRARY

**Would it smell as sweet by any other name?**

had a name — an occurrence almost as old as taxonomy itself. The result is that the same plant can have many names, and not everyone knows which one to use. Such synonyms are a particular problem in the study of medicinal plants, says Alan Paton, a plant taxonomist and bioinformatician at Kew Gardens in London.

The TNRS was built with financial and technical support from iPlant, a project run by the US National Science Foundation to fund cyberinfrastructure for plant science. It corrects names by comparing lists that users feed into it with the 1.2 million names in the Missouri Botanical Garden's Tropicos database, one of the most authoritative botanical databases. If the TNRS cannot find a name in Tropicos, it uses a fuzzy-matching algorithm, similar to a word-processor's spellchecker, to find and correct misspellings. It also hunts through Tropicos's lists of alternative names and supplies the one that is most up to date. When Enquist ran the 611,728 names through the system, just 202,252 came back, showing that two-thirds of them were invalid.

Because Tropicos is less comprehensive for plants outside the Americas, the team hopes to link the TNRS with The Plant List ([www.theplantlist.org](http://www.theplantlist.org)), a collaborative compilation of databases from Kew and other sources. Launched online in December 2010, it aims to become a global record of plants. The scientists are also working on a tool to correct geographical data — one that knows, for example, that Brazil, Brasil and Brésil are the same place, and can recognize when someone has muddled up longitude and latitude. ■

Each bespoke knockout in the Sanger group's library contains an added 'conditional allele'. This allows scientists to disrupt gene function in a living mouse at any body site and at any point in the animal's development by the timely addition of enzymes that recognize the inserted allele. By this means, the effects of the missing gene do not kill the mouse before the researchers have a chance to study it.

"It is truly a feat of genius," says Geoff Hicks, a geneticist at the University of Manitoba in Winnipeg who leads the Canadian contribution to the IKMC. "This paper really pushed the technology in an extremely innovative way and met a challenge that seemed unattainable."

Various groups in the international effort are using other, non-conditional techniques to inactivate thousands more genes. Researchers in Texas, Canada and Germany have mutated close to 12,000 genes using an untargeted approach called gene trapping, and Regeneron Pharmaceuticals, a company based in Tarrytown, New York, has specifically targeted around 3,500 genes using a technology that works well in smaller genes but results in mice that are less flexible for research than conditional knockouts. "The approaches are complementary," says Aris Economides, Regeneron's senior director of genome engineering technologies. "This is going to play out well for the end user."

To date, nearly 17,000 different genes have been knocked out, leaving only around 3,000 more to go. The Sanger team, however, hopes to replace most of the genes hit by gene trapping with conditionally targeted knockouts, because targeting allows individual genes to be manipulated with greater precision.

Already, mutant mice have been generated from almost 1,000 of the embryonic stem-cell lines obtained, and the IKMC repositories in the United States, Canada and Europe receive hundreds of new orders every month. The next challenge is to study the function of each missing gene. To this end, the US National Institutes of Health last year committed \$110 million over the next five years to characterize around 2,500 of the IKMC's mutant mice through the International Mouse Phenotyping Consortium, with plans for another \$110 million to define 5,000 more if the first phase is successful.

"Knocking out the mice is simple relative to the huge task of finding out what all those genes do," says Richard Finnell, a geneticist at the Texas A&M Health Science Center in Houston. ■

**"It is one of the most significant biological resources in the past century of science."**

## BOTANY

# Species spellchecker fixes plant glitches

*Online tool should weed out misspellings and duplications.*

BY JOHN WHITFIELD

Brian Enquist and his collaborators were delighted with their freshly compiled data set of 22.5 million records on the distribution and traits of plants in the Americas. But their delight turned to horror when they realized that the data set contained 611,728 names: nearly twice as many as there are thought to be plant species on Earth.

Completed in December 2010, the records were intended to help Enquist and his colleagues to discern trends in how forest trees in a wide variety of environments respond to climate change. But the data were clearly full of bogus names, making it impossible to count the species in a particular area, or their relative abundance. "I started to question our ability even to compare something as basic as species diversity at two sites," says Enquist, a plant ecologist at the University of Arizona in Tucson.

This month, Enquist's team will unveil a solution that could help botanists and ecologists worldwide. The Taxonomic Names Resolution Service (TNRS) aims to find and fix the incorrect plant names that plague scientists' records.

"It looks really good," says Gabriela Lopez-Gonzalez, a plant ecologist at the University of Leeds, UK, who curates a database of forest plots. Fixing species lists by hand is arduous, she says. "This should save us a lot of time".

She and others agree that the problem is widespread in botanical databases. "Digitization has made the problem worse," says TNRS co-leader, botanist Brad Boyle, also at the University of Arizona. Boyle explains that as more data are added to digital records, the chance of introducing errors also increases. Even in herbarium specimens, which ought to be the gold standard for plant identification, about 15% of the names are misspelt, he says.

Many of the errors seem to arise because biologists are not as careful as they should be when entering data into digital records. The TNRS team estimates that about one-third of the names entered into online repositories — such as GenBank, the US National Institutes of Health collection of DNA-sequence data, or the Ecological Society of America's VegBank database of plant-plot data — are incorrect.

The other problem is that names change. Old names can be abolished when experts reclassify plants as ideas about evolutionary relationships change, or when they realize the species already



R. EVANS/PHOTOLIBRARY

**Would it smell as sweet by any other name?**

had a name — an occurrence almost as old as taxonomy itself. The result is that the same plant can have many names, and not everyone knows which one to use. Such synonyms are a particular problem in the study of medicinal plants, says Alan Paton, a plant taxonomist and bioinformatician at Kew Gardens in London.

The TNRS was built with financial and technical support from iPlant, a project run by the US National Science Foundation to fund cyberinfrastructure for plant science. It corrects names by comparing lists that users feed into it with the 1.2 million names in the Missouri Botanical Garden's Tropicos database, one of the most authoritative botanical databases. If the TNRS cannot find a name in Tropicos, it uses a fuzzy-matching algorithm, similar to a word-processor's spellchecker, to find and correct misspellings. It also hunts through Tropicos's lists of alternative names and supplies the one that is most up to date. When Enquist ran the 611,728 names through the system, just 202,252 came back, showing that two-thirds of them were invalid.

Because Tropicos is less comprehensive for plants outside the Americas, the team hopes to link the TNRS with The Plant List ([www.theplantlist.org](http://www.theplantlist.org)), a collaborative compilation of databases from Kew and other sources. Launched online in December 2010, it aims to become a global record of plants. The scientists are also working on a tool to correct geographical data — one that knows, for example, that Brazil, Brasil and Brésil are the same place, and can recognize when someone has muddled up longitude and latitude. ■



## MISCONDUCT

# Europe tackles huge fraud

*Regulators scramble to recover millions of euros awarded to fake research projects.*

BY QUIRIN SCHIERMEIER

Stifling bureaucracy is often blamed for discouraging scientists and businesses from participating in the research programmes of the European Commission (EC). But the commission's notoriously cumbersome procedures and rigid control mechanisms have apparently not prevented a criminal syndicate from conducting a brazen fraud that has siphoned off millions in EC grant funds.

Italian authorities and the European Anti-Fraud Office (OLAF) in Brussels, Belgium, have confirmed that they are prosecuting members of a large network accused of pocketing more than €50 million (US\$72 million) in EC grants for fake research projects. In Milan, Italy, the Finance Police last month charged several individuals in relation to the fraud. In Brussels, meanwhile, the EC has terminated four collaborative projects in information technology, and excluded more than 30 grant-winners from participation in around 20 ongoing projects. Investigations are still under way in the United Kingdom, France, Greece, Austria, Sweden, Slovenia and Poland.

"We don't have any records of [previous] fraud at such a scale," says David Boubilil, the commission's spokesman for taxation, customs, anti-fraud and audit. While investigations continue, Italian prosecutors and OLAF will not disclose the names of the suspects, or the research projects with which they were involved.

The fraud has been conducted in a "highly sophisticated manner, resembling money laundering", by means of a cross-border network of fictitious companies and subcontractors, says Pavel Bořkovec, a spokesman for OLAF. Several project coordinators stand accused of

having claimed inflated costs, or expenses for non-existent research activities and services, he says.

"The projects were apparently organized with the sole intention to deceive the commission and its control mechanisms," says Boubilil. To make them seem legitimate, grant applications included the names of real

research projects. But the commission does extensive checks on project partners, including companies, which are meant to catch large-scale fraud. The success of the fraud suggests that those involved were unusually familiar with weaknesses in the EC's procedures, and adept at forging legal documents.

Boubilil insists that the commission has learned lessons from the case. All departments handling research grants — including the EC's Information Society and Media Directorate General, which oversaw the terminated projects — are now trained to look out for the methods used by the network. Guidelines for evaluating projects and their partners are set to be updated. The EC has already recovered €10 million of the money, and will seek to recover the rest through the courts, Boubilil says.

The commission is currently developing a multibillion-dollar 'Common Strategic Framework' which, from 2014, will combine its various funding streams into a single channel for all research and innovation funding. Concerned about the burden of Brussels bureaucracy, several thousand European scientists signed a petition this year ([www.trust-researchers.eu](http://www.trust-researchers.eu)) calling for the framework to be "based on mutual trust and responsible partnering". Some now fear that the fraud could hamper efforts to cut red tape.

"I'm worried that some will argue that what has happened proves that we need more rather than less control," says

Herbert Reul, chair of the European Parliament's committee on industry, research and energy, which supports the simplification of the EC's funding procedures. "I sincerely hope that this will not happen. Actually, it is a good sign that this worrying attempt at deceiving the commission has been discovered and will be punished." ■



Herbert Reul, the European Parliament research committee chair, hopes a case of fraud will not hinder easing of grant applications.

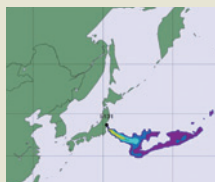
scientists, established research institutes and existing companies, he says. But in most cases the alleged project partners were included without their knowing.

Insiders in Brussels say that rare cases of minor financial dishonesty, from inflated invoices to smaller cases of embezzlement, are regarded as unavoidable in large collaborative

EU 2011 PE-EP

  
**MORE  
ONLINE**

#### TOP STORY



Researchers call for release of nuclear data  
[go.nature.com/s18k6k](http://go.nature.com/s18k6k)

#### OTHER NEWS

- British Council axes climate-change work [go.nature.com/ht7ahj](http://go.nature.com/ht7ahj)
- The worrying dearth of research into neuropsychopharmacology [go.nature.com/arb8xf](http://go.nature.com/arb8xf)
- How dinosaur collagen survived for millions of years [go.nature.com/iy9c3f](http://go.nature.com/iy9c3f)

#### PODCAST



Voyager 1 reports back, and the rap guide to evolution  
[go.nature.com/qeajjm](http://go.nature.com/qeajjm)

ZAMG





Students will be grateful for a substantial boost to Egypt's education budget.

MIDDLE EAST

# Egypt invests in its science

*Latest budget establishes research as a national priority.*

BY DECLAN BUTLER

Four months after Egypt's revolution toppled the authoritarian regime of President Hosni Mubarak, science and education are slowly emerging from the post-revolution chaos as national priorities. Revitalizing Egypt's sclerotic and chronically underfunded research, education and innovation systems will require sweeping reforms and substantial rises in spending. But modest funding increases and a warmer political climate for research and education have left Egyptian scientists feeling more optimistic than ever before.

"We are going to build our economy to be based on democracy, and science and technology," says Maged Al-Sherbiny, president of the Academy of Scientific Research and Technology in Cairo and assistant minister for research.

➔ **NATURE.COM**  
For more on the Arab awakening, see:  
[go.nature.com/ox5oag](http://go.nature.com/ox5oag)

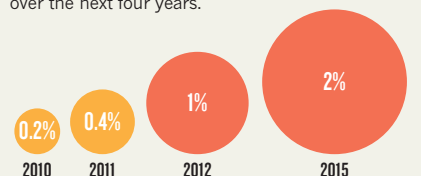
On 1 June, the Egyptian cabinet approved the first post-revolution budget, which boosted

science despite the severe social and economic crises gripping the country. Research spending will rise from E£2.4 billion (US\$404 million) to E£3 billion in the 2011–12 financial year. The education budget also jumped, by 16% to E£55.7 billion.

The increase in science spending still leaves it at only around 0.4% of gross domestic product (GDP), much less than the 1–2% that researchers say they would like. The goal is to reach that level within 4 years, says Al-Sherbiny (see 'Grand plans'). "That target is optimistic," cautions Tarek Khalil, president

## GRAND PLANS

Egypt is planning to substantially increase the proportion of GDP that it spends on research over the next four years.



and provost of Nile University in Cairo, "but if we can do it, great."

Over the next three years, the government plans to create 50,000 research posts for young researchers, most of which will be government jobs at universities and research institutes. An extra E£2 billion has been secured for the plan, according to Al-Sherbiny. Several thousand of the posts would be subsidized positions within industry, part of a broader goal to boost the almost non-existent levels of research in the private sector, he adds.

Al-Sherbiny and science minister Amr Ezzat Salama have also proposed other reforms, including raising researchers' salaries and introducing performance-based bonuses, for which E£1.3 billion has been secured, says Al-Sherbiny. Under the proposed reforms, the number of government research institutes would increase from 198 to 258 — including large new centres in microelectronic systems and solar energy. The expanded research efforts will focus on seven areas considered important for Egypt: renewable energy, with an emphasis on solar and wind; water, including desalination, irrigation and groundwater management; food and agriculture; health, including hepatitis C, cancer and obesity; information technology; space; and socioeconomic goals such as increasing science in the classroom.

The fall of Mubarak may also finally open the way to a decade-old proposal by Nobel laureate Ahmed Zewail, an Egyptian-born chemist at the California Institute of Technology in Pasadena, to create a US\$2-billion independent, non-profit science city that would include centres of excellence, hire top researchers and teach the cream of the country's students. The state has provided 120 hectares in 6th of October City, outside Cairo, but public funding for the project looks set to be minimal. Zewail still needs to raise most of the \$1 billion needed to establish the city, and a further \$1 billion as an endowment, through philanthropy and foreign aid. Zewail has set up a board of trustees that includes six Nobel laureates and other prominent individuals such as Susan Hockfield, president of the Massachusetts Institute of Technology in Cambridge, who will direct the project.

Farouk El-Baz, an Egyptian-born geologist at Boston University in Massachusetts, says that even though the political interest in science has not yet translated into adequate funding and reforms, scientists must take into account the many other pressing post-revolution demands, and be patient. "I don't think the reforms are enough yet, but they are going in the right direction; there is no question about that."

Al-Sherbiny says that solidarity from the international scientific community will help. "This is a time when our friends and partners need to stand by us to help us realize our dreams, to offer to work together, to offer expertise and money, to help us build the new system we are trying to establish," he says. ■

C. WIENS/ALAMY

SOURCE: M. AL-SHERBINY



Chimpanzees at the New Iberia Research Center in Louisiana are some of very few remaining worldwide that are still being used in invasive research.

# CHIMPANZEE RESEARCH ON TRIAL

*As pressure from activists builds, the United States is considering whether it should end invasive experiments in chimpanzees.*

BY MEREDITH WADMAN

**T**he unusual meeting was held in a conference room, but it might have been called a war room. Gathered inside a little-known research centre in southern Louisiana, the people who oversee chimpanzee research in the United States were preparing to battle for the survival of their enterprise.

Although no other country besides Gabon carries out invasive experiments with chimpanzees, the United States continues such work at three major research facilities. Louisiana's New Iberia Research Center (NIRC) is the largest, with a population of 360 chimps, used by investigators from pharmaceutical companies and federal agencies to test new drugs and study diseases such as hepatitis. During the meeting, Thomas Rowell, director of the NIRC, stood up, surveyed the audience, and launched into a presentation about possible strategies to build public support for their work.

"How do we get industry to be forthcoming about their use of chimpanzees?" a slide read.

"Could we get at least a few solid examples of how the use of chimpanzees has truncated the time to discovery?"

And "When we talk about time and lives saved by using chimpanzees,

can we provide actual time span data or numbers?"

Another slide went on to note that the National Institutes of Health (NIH) spends about US\$12 million a year caring for the chimpanzees it supports (currently totalling 734), versus the billions in health-care costs for the human diseases that can be studied through experiments on chimpanzees. One of them, hepatitis C, currently affects at least 170 million people globally. If researchers don't have access to the chimp model, said Rowell, people afflicted with hepatitis C will suffer. "Their lifespans are going to be shortened. They will not have a proper quality of life." He called them a "silent voice".

Rowell's pep talk in April was partly for the benefit of some visitors at the meeting: representatives from the Food and Drug Administration, the National Institute of Allergy and Infectious Diseases, the drug industry and, most importantly, the Institute of Medicine (IOM). The IOM, the medical branch of the independent National Academy of Sciences, was asked by the NIH in January to examine whether the government should keep supporting biomedical research on chimpanzees — the closest living relatives of *Homo sapiens*.

The NIH called for the study after the agency sparked a storm of

J. BREAUX



opposition last year, when it announced plans to move 186 semi-retired chimps back into active research<sup>1</sup>. After protests by the Humane Society of the United States (HSUS) in Washington DC, famed primatologist Jane Goodall and others, the NIH changed course and said that it would make no decision on moving the chimps until the IOM study is complete. The study, it announced, would be “an in-depth analysis to reassess the scientific need for the continued use of chimpanzees to accelerate biomedical discoveries”.

Proponents say that the research is necessary for continued progress towards a hepatitis-C vaccine; for developing more effective drugs against hepatitis B and C; for testing monoclonal antibody treatments for a variety of conditions; and for research to develop a vaccine against respiratory syncytial virus, a seasonal virus that kills more than more than 66,000 children under the age of 5 each year across the globe<sup>2</sup>. For many of these conditions, backers argue, the chimpanzee is either the only available model, or by far the best one.

But chimpanzee research in the United States is facing growing public and political opposition. Animal-welfare activists have stepped up their efforts to end the work, arguing that it is inhumane, ineffective and a waste of taxpayer money. The day after the meeting, activists held a press conference on Capitol Hill to mark the introduction of the Great Ape Protection and Cost Savings Act. The act would make all invasive chimpanzee research illegal, including private-sector work conducted at the centres and paid for by drug companies. The bill's lead sponsor in the House of Representatives is Roscoe Bartlett (Republican, Maryland), who trained as a physiologist and conducted primate research with NASA and with the military in the 1960s.

“There’s just no valid argument to continue to keep these great apes as they’re now being kept,” Bartlett told the news conference. “Very few of them are used in research and I’m not sure that any of them need to be used.”

The scrutiny this year adds to the tension felt by researchers who work with chimpanzees. That stress is particularly intense at the NIRC, which has been on the defensive ever since a television documentary two years ago showed footage of employees there mistreating and neglecting chimpanzees and macaques. The NIRC, which is part of the University of Louisiana at Lafayette, later paid a fine and has since passed numerous inspections, but the exposé helped to propel the activism. In the contest for public support, says Rowell, “our backs are up against the wall”.

## A STUDY UP CLOSE

On the same day that the chimpanzee-protection measure was introduced in Congress, staff at the NIRC prepared to start a drug-company trial that used two chimpanzees to test the absorption, metabolism and excretion of an experimental medication. One of the animals was Simba, an 88-kilogram male around 40 years old. That morning he was coaxed

from his outdoor enclosure, where he lives in a large social group, into an individual cage. A technician used a needle and syringe to sedate him. He was then strapped to a stretcher and transported by ambulance to Building 52 to receive a pre-study physical examination.

At 9:27 a.m., Simba was slid off the stretcher — where it became clear that he had defecated — and onto a stainless-steel gurney. His fleshy pink gums were relaxed and prominent. He was drooling.

“I need to do a dental on him,” said Dana Hasselschwert, head of the veterinary-sciences division and one of nine veterinarians on the NIRC staff. The veterinarians care for the centre’s chimps, along with its 6,500 macaques and other monkeys. Today, three technicians are assisting Hasselschwert with the physical. Speed is important, because the sedative is short-lived. Fully alert chimpanzees are strong and sometimes violent.

One technician quickly shaved Simba’s forearms, armpits and groin. On the skin of his right groin, a tattoo identified him as chimpanzee number xo19. The other technicians placed electrodes on his body;

his electrocardiogram revealed a regular heart rhythm. Simba’s blood pressure was 143/87 millimetres of mercury — normal for him, Hasselschwert said. Blood was drawn from Simba’s left femoral vein; his rectal temperature was taken and was normal, at 37.3 °C. His pulse was 104 beats per minute; his respirations 32.

Hasselschwert palpated his liver and kidneys and found nothing abnormal. But one of the technicians was having trouble

catheterizing Simba to collect a urine sample. Hasselschwert placed an ultrasound paddle on Simba’s lower abdomen and located his bladder on a nearby screen. An assistant quickly shaved the overlying area.

“It’s undignified, a male having bikini marks,” Hasselschwert declared. She inserted a needle through Simba’s abdominal wall and withdrew three millilitres of pale yellow urine.

Simba’s breathing was speeding up, a sign of growing wakefulness. “Y’all, we need to move,” Hasselschwert said. She wiped Simba’s drooling gums with paper towels, and patted his open palm. His hand was half again as big as hers. “He looks good,” she declared, and, at 9:40 a.m., Simba was wheeled away on the gurney and placed in a wire cage that measured 2 metres long by 1.5 metres wide by 2.2 metres high. The cage is one of many in the room, and it can be compressed if an animal refuses to present an appendage for injections or blood withdrawal — a procedure that staff call “squeezing up”. Three days later, Simba would be injected with the experimental drug. After that, for 72 hours, at regular intervals, his blood would be drawn and his urine collected from a pan beneath the cage. He would then be returned to his outdoor enclosure.

Last year, the NIRC conducted 23 chimpanzee studies, which typically involve between two and six animals. On the day of Simba’s physical, ten chimps were in experiments. The remaining chimps are kept in the outdoor cages. To keep the chimps prepared for being research subjects, trainers reward them with fruit in exchange for presenting their legs for mock injections, or for urinating in a cup. The chimps are wary of strangers, at whom they are wont to hurl gravel or faeces.

Chimpanzee studies are expensive, costing anywhere from \$20,000 to \$250,000. And roughly 85% of the revenue for the NIRC comes from a score of pharmaceutical companies that are regular customers. (Other centres tilt more towards academic and governmental clients.) As well as conducting drug and vaccine trials, the NIRC breeds macaques for several companies, and is a registered importer of the monkeys.

The other 15% of the centre’s revenue comes from government agencies, mainly the NIH. The biomedical agency owns 124, or roughly one-third, of the NIRC chimpanzees, and pays the centre to maintain two breeding colonies of macaques. The centre also conducts chimpanzee research under contract for the Centers for Disease Control and Prevention. It owns 11 more chimps, which are kept at Bioqual, a company in Rockville, Maryland, where young animals are used in

**“IT IS UNETHICAL NOT TO USE THE CHIMP MODEL FOR CERTAIN INDICATIONS.”**

S. BREAU



Thomas Rowell directs the New Iberia Research Center in Louisiana.



hepatitis-C studies run by the NIH's infectious-diseases institute.

If the IOM were to recommend that the NIH stop supporting chimpanzee research, and if the NIH were to comply, this would, theoretically, not affect the drug-company funded research at the NIRC and the other centres. But in practice, the directors say, it would hobble their enterprise, not least because some two-thirds of the chimpanzees available for research in the United States are owned or supported by the NIH (see 'Chimpanzee Research in the United States'). What is more, they say, the per diem fees and user fees paid by companies for individual experiments do not begin to cover the long-term care of the animals, which is supported by NIH infrastructure grants.

"The lifetime maintenance of chimpanzees requires a long-term commitment of financial support that individually sponsored studies do not provide," the directors wrote in a jointly authored statement to *Nature*.

## INSPIRED WORK

Rowell, who is 52, has been working with chimps most of his life, ever since he took a job cleaning cages at the NIRC when he was 17. He quickly got a taste of the value of chimpanzee research, when Carleton Gajdusek shared the 1976 Nobel Prize in Physiology or Medicine for discovering that neurodegenerative disorders such as kuru and scrapie are transmitted by infectious agents<sup>3</sup>. As part of his research, Gajdusek injected infected human brain tissue into chimps from the centre.

The thrill of Gajdusek's work rubbed off on Rowell. "This is what was so exciting — a teenager off the street working at the level that I was — and being involved with something so huge." Rowell chose his career at that point, and earned a degree in veterinary medicine at the Louisiana State University School of Veterinary Medicine. Hired by the NIRC in 1990, he became its director in 1998.

Rowell has expanded the centre significantly, from about 170 employees in 1998 to 249 today, and from 4,560 primates in 1998 to 6,860 today. He also strengthened the NIRC's experimental credentials and abilities, which has made the centre highly attractive to the pharmaceutical industry.

In the years since Rowell first started working at the facility, support for chimp research has slowly eroded around the world. The United States stopped importing chimpanzees after signing a 1973 treaty banning trade in endangered species. When the AIDS epidemic hit, the NIH launched a breeding programme for chimpanzees, but the agency declared a moratorium on breeding in 1995, after it became clear that chimps were a poor model for the disease.

Soon, countries started to outlaw chimp research completely. In 1997, the United Kingdom took that step. Another eight countries followed suit in the next decade, and last year, the European Union outlawed great-ape experimentation.

Only one pharmaceutical company, GlaxoSmithKline, has dropped chimp research, at least publicly. It announced in 2008 that "the case for using great apes in the future is less clear than it may have been previously".

Opponents of chimp research have painted the United States as an outlier for continuing to allow such experiments. That charge irks the directors of the chimpanzee centres. Responding to a request from *Nature*, the directors catalogued 27 chimpanzee studies carried out at their centres by foreign companies or scientists since 2005.

"The Europeans did not ban their companies from coming to the United States," says John VandeBerg, director of the Southwest National Primate Research Centre in San Antonio, Texas, another of the centres that conducts chimp research. "And I can assure you they are not going to ban the importation of drugs into their countries that are developed using chimpanzees."

In the United States, public pressure to shut down the research intensified after the television exposé. The show contained video

footage obtained surreptitiously by an HSUS investigator who infiltrated the NIRC and worked there for nine months.

In one scene from the resulting documentary, aired by the ABC news show *Nightline*, a sedated chimp fell several feet from a bench to the cement floor of a cage. In another scene, a sedated chimp is carried by its arms and legs, not on a stretcher.

Footage of some of the centre's monkeys was equally damaging. A technician hit a baby on the head after it bit her, and another employee rapped a monkey's teeth with a metal pole. In a different scene, an anaesthetized monkey was allowed to fall from a chest-level counter to the floor of a lab room.

The show drew strong reactions. Jane Goodall issued a statement saying: "In no lab I have visited have I seen so many chimpanzees exhibit such intense fear." And agriculture secretary Tom Vilsack ordered an investigation of the NIRC's animal-welfare practices. In the following 14 months, the centre underwent inspections by two units of the US Department of Agriculture (USDA), the NIH's Office of Laboratory Animal Welfare in Bethesda, Maryland, the Association for Assessment and Accreditation of Laboratory Animal Care International in Frederick, Maryland, and auditors from every one of the NIRC's pharmaceutical clients. The government agencies also paid surprise visits to the other facilities conducting chimpanzee research.

## THE CASE FOR RESEARCH

In May 2010, the NIRC paid \$18,000 to the USDA to settle six alleged violations of the Animal Welfare Act, such as leaving sedated adult chimps unattended with nursing infants. As part of the agreement, the centre neither admitted nor denied that the violations took place, or that they were, in fact, violations. But the NIRC has since retrained employees in, for instance, keeping animals safe when they are sedated.

Rowell says that he watched all ten hours of undercover video footage that the HSUS turned over to the NIH. He concedes that there were moments of carelessness and one case of inappropriate behaviour, when the technician hit the infant monkey. But he says that the undercover operative — who was working as an aide — bears responsibility for the fall involving the sedated monkey. It was her job to protect the animal, but she had stepped away to film the room from a distance. Overall, he says, "I was proud of what I saw".

Now, focused by the IOM study, he is going on the offensive. Others are also speaking up, such as Christopher Walker, a hepatitis-C researcher based at Nationwide Children's Hospital in Columbus, Ohio, who is the main academic customer at the NIRC. Walker is part of a team funded by a five-year, \$12.5-million grant from the Bill & Melinda Gates Foundation in Seattle, Washington, to try to develop a drug for hepatitis C that reinvigorates exhausted immune-system cells called T cells; he also relies on NIH grants.

Walker has not spoken to the press before about his work with

**"STOP USING THE  
EXCUSE THAT  
CHIMPS ARE  
ESSENTIAL TO  
THIS RESEARCH."**

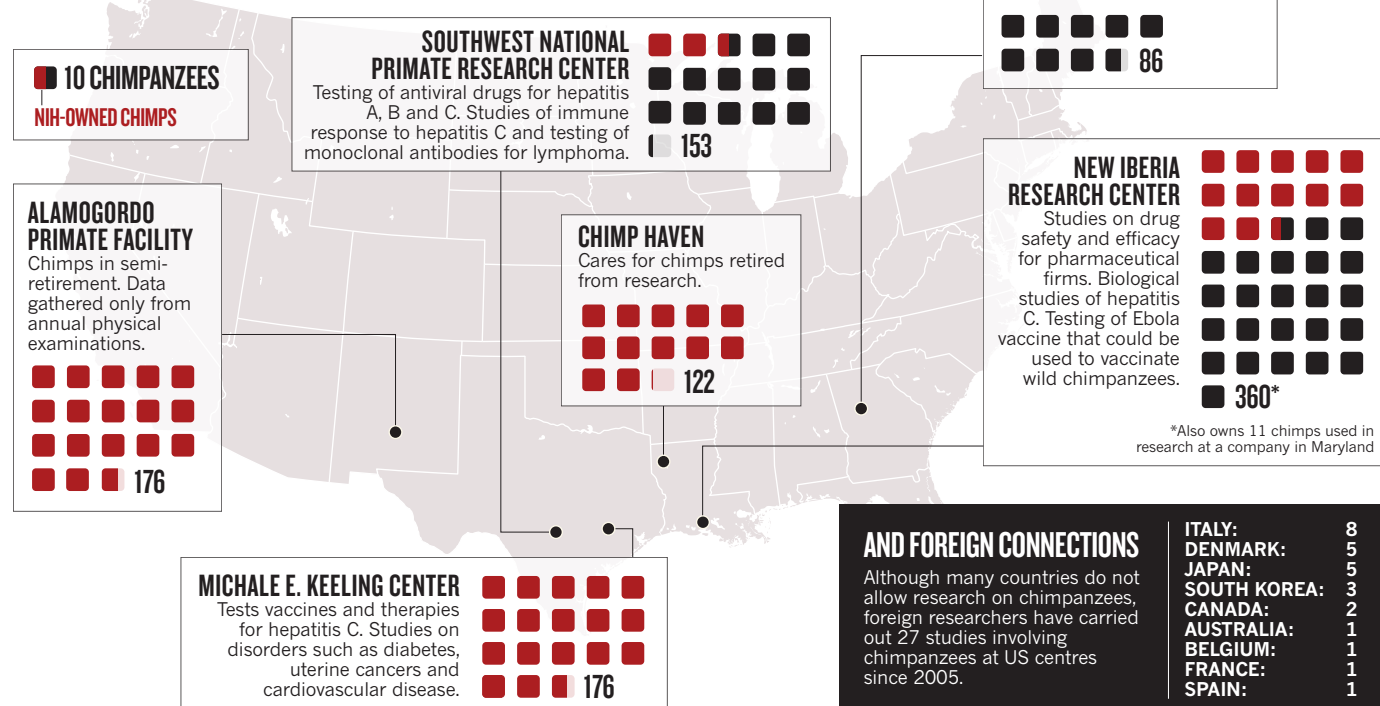


Outdoor enclosures at the NIRC house chimps in social groups.

J. BREAUX

## CHIMPANZEE RESEARCH IN THE UNITED STATES

This year, the US National Institutes of Health (NIH) will spend US\$12 million maintaining chimps for research and providing care for retired animals. Three centres conduct invasive research, one does behavioural work and two care for chimps not currently involved in research. The \$12 million does not include extramural grants.



chimpanzees; he has been afraid of being targeted by animal-rights activists. But he is talking now, he says, “because we are reaching a critical decision point”. Walker’s work focuses on unravelling the role of cellular immunity in hepatitis-C infection, which often leads to liver cancer, a disease that is almost always fatal without a liver transplant. While working at a firm called Chiron in Emeryville, California, in the 1990s, Walker did the scientific groundwork<sup>4,5</sup> in chimpanzees that led Merck to develop a hepatitis-C vaccine programme. Several Merck employees then spun off Okairos, a Rome-based biotech, which has since moved a vaccine into human trials, after publishing proof-of-concept work in chimpanzees<sup>6</sup>.

“The chimpanzees were absolutely critical,” Walker says, in establishing that immune-system cells called T-cells have an important role in controlling the hepatitis-C virus, and that any successful vaccine would need to generate a T-cell response.

Walker is a strong believer that chimpanzee studies continue to be needed not only for developing a hepatitis-C vaccine, but also for testing the safety of new, and potentially risky, medicines to treat both hepatitis C and B. He points, for instance, to research published in 2009 that showed RNA silencing to be effective in controlling hepatitis-C infection in chimps<sup>7</sup>.

Some others who use chimpanzees see few remaining justifications for experiments on the animals. Michael Houghton, a virologist at the University of Alberta in Edmonton, and a co-discoverer of the hepatitis-C virus, says that in research related to that virus, “we do not need the chimp any more for diagnostic development or for antiviral-drug development as we have the infected human available”. The risk-to-benefit ratio for infected people in such studies is low enough to justify testing in humans, he says.

Still, Houghton supports chimpanzee use for hepatitis-C vaccine development, because vaccines must be tested in uninfected individuals. He also supports studies in chimpanzees for the development of riskier immunotherapies against the disease. “As inconveniencing tens of chimpanzees impacts the health

of millions of humans, it is unethical not to use the chimp model for certain indications,” says Houghton. He also believes that it would be unwise not to keep humanely treated chimps available in sanctuaries in case of bioterrorist attacks; the animals could be used to study the transmission of infectious bioweapons as well as vaccines and therapies, he says.

Activists, though, see no rationale for continuing tests on chimps, partly, they say, because ever-more sophisticated *in vitro* methods make it unnecessary<sup>8</sup>. They also argue that, despite the genetic similarities between chimps and humans, they have relevant differences in, for instance, immune-response genes<sup>9</sup>, and that differences in gene expression make chimps weak as a biological model. “Stop using the excuse that chimps are essential to this research,” says John Pippin, a physician who is senior medical and research adviser for the Physicians Committee for Responsible Medicine in Washington DC.

By the end of the year, the IOM committee will offer its own analysis of whether chimp research is scientifically warranted. The committee’s report will be the most weighty pronouncement on the issue so far in the United States, but it may not settle the debate.

The ongoing controversy has taken a toll on some of those who work with chimpanzees. Rowell says he does not take the same amount of pleasure in his work that he did five years ago. “I’m exhausted,” he says. Still, he vows to stay in the job. “It’s not something that I do. It’s who I am.” ■ **SEE EDITORIAL P.251**

**Meredith Wadman** is a reporter for Nature based in Washington DC.

1. *Nature* **467**, 507–508 (2010).
2. Nair, H. *et al. Lancet* **375**, 1545–1555 (2010).
3. Gajdus, D. C., Gibbs, C. J. & Alpers, M. *Nature* **209**, 794–796 (1966).
4. Grakoui, A. *et al. Science* **302**, 659–662 (2003).
5. Shoukry, N. H. *et al. J. Exp. Med.* **197**, 1645–1655 (2003).
6. Folgori, A. *et al. Nature Med.* **12**, 190–197 (2006).
7. Lanford, R. E. *et al. Science* **327**, 198–201 (2010).
8. Masaki, T. *et al. J. Virol.* **84**, 5824–5835 (2010).
9. Bettauer, R. H. *J. Med. Primatol.* **39**, 9–23 (2010).

**NATURE.COM**  
For more on chimp science, see:  
[go.nature.com/5uusy2](http://go.nature.com/5uusy2)





# The dawn of quantum biology

*The key to practical quantum computing and high-efficiency solar cells may lie in the messy green world outside the physics lab.*

BY PHILIP BALL

On the face of it, quantum effects and living organisms seem to occupy utterly different realms. The former are usually observed only on the nanometre scale, surrounded by hard vacuum, ultra-low temperatures and a tightly controlled laboratory environment. The latter inhabit a macroscopic world that is warm, messy and anything but controlled. A quantum phenomenon such as 'coherence', in which the wave patterns of every part of a system stay in step, wouldn't last a microsecond in the tumultuous realm of the cell.

Or so everyone thought. But discoveries in recent years suggest that nature knows a few tricks that physicists don't: coherent quantum processes may well be ubiquitous in the natural world. Known or suspected examples range from the ability of birds to navigate using Earth's magnetic field to the inner workings of photosynthesis — the process by

DAREN NEWMAN

which plants and bacteria turn sunlight, carbon dioxide and water into organic matter, and arguably the most important biochemical reaction on Earth.

Biology has a knack for using what works, says Seth Lloyd, a physicist at the Massachusetts Institute of Technology in Cambridge. And if that means “quantum hanky-panky”, he says, “then quantum hanky-panky it is”. Some researchers have even begun to talk of an emerging discipline called quantum biology, arguing that quantum effects are a vital, if rare, ingredient of the way nature works. And laboratory physicists interested in practical technology are paying close attention. “We hope to be able to learn from the quantum proficiency of these biological systems,” says Lloyd. A better understanding of how quantum effects are maintained in living organisms could help researchers to achieve the elusive goal of quantum computation, he says. “Or perhaps we can make better energy-storage devices or better organic solar cells.”

## ENERGY ROUTEFINDER

Researchers have long suspected that something unusual is afoot in photosynthesis. Particles of light called photons, streaming down from the Sun, arrive randomly at the chlorophyll molecules and other light-absorbing ‘antenna’ pigments that cluster inside the cells of every leaf, and within every photosynthetic bacterium. But once the photons’ energy is deposited, it doesn’t stay random. Somehow, it gets channelled into a steady flow towards the cell’s photosynthetic reaction centre, which can then use it at maximum efficiency to convert carbon dioxide into sugars.

Since the 1930s, scientists have recognized that this journey must be described by quantum mechanics, which holds that particles such as electrons will often act like waves. Photons hitting an antenna molecule will kick up ripples of energized electrons — excitons — like a rock splashing water from a puddle. These excitons then pass from one molecule to the next until they reach the reaction centre. But is their path made up of random, undirected hops, as researchers initially assumed? Or could their motion be more organized? Some modern researchers have pointed out that the excitons could be coherent, with their waves extending to more than one molecule while staying in step and reinforcing one another.

If so, there is a striking corollary. Coherent quantum waves can exist in two or more states at the same time, so coherent excitons would be able to move through the forest of antenna molecules by two or more routes at once. In fact, they could simultaneously explore a multitude of possible options, and automatically select the most efficient path to the reaction centre.

Four years ago, two teams working under Graham Fleming, a chemist at the University of California, Berkeley, were able to obtain experimental proof to back up this hypothesis (See ‘Quantum fact meets fiction’). One team used a string of very short laser pulses to probe the photosynthetic apparatus of the green sulphur bacterium *Chlorobium tepidum*<sup>1</sup>. The researchers had to chill their samples to 77 K with liquid nitrogen, but the data from their laser probes showed clear evidence of coherent exciton states. The second team carried out a similar study of the purple bacterium *Rhodobacter sphaeroides*<sup>2</sup>, and found much the same electronic coherence operating at temperatures up to 180 K.

In 2010, researchers from the first group published evidence of quantum coherence in their bacterial complex at ambient temperatures<sup>3</sup> — showing that coherence is not just an artefact of cryogenic laboratory conditions, but might actually be important to photosynthesis in the real world. Around the same time, a team led by Gregory Scholes, a chemist at the University of Toronto in Canada, also reported coherence effects at ambient temperatures<sup>4</sup> — this time not in bacteria, but in photosynthetic cryptophyte algae, evolutionarily distinct organisms that are more closely related to plants and animals, and that use completely

different light-absorbing chemical groups.

But how can quantum coherence last long enough to be useful in photosynthesis? Most physicists would have assumed that, at ambient temperatures, the surrounding molecular chaos in the cell destroys the coherence almost instantly.

Computer simulations carried out by Lloyd and some of his colleagues suggest an answer: random noise in the environment might actually enhance the efficiency of the energy transfer in photosynthesis rather than degrade it<sup>5</sup>. It turns out that an exciton can sometimes get trapped on particular sites in the photosynthetic chain, but simulations suggest that environmental noise can shake it loose gently enough to avoid destroying its coherence. In effect, says Lloyd, “the environment frees up the exciton and allows it to get to where it’s going”.

Photosynthesis is not the only example of quantum effects in nature. For instance, researchers have known for several years that in some enzyme-catalysed reactions<sup>6</sup>, protons move from one molecule to another by the quantum-mechanical phenomenon of tunnelling, in which a particle passes through an energy barrier rather than having

## Nature knows a few tricks that physicists don’t.

to muster the energy to climb over it. And a controversial theory of olfaction claims that smell comes from the biochemical sensing of molecular vibrations — a process that involves

electron tunnelling between the molecule responsible for the odour and the receptor where it binds in the nose<sup>7</sup>.

Are such examples widespread enough to justify a whole new discipline, though? Robert Blankenship, a photosynthesis researcher at Washington University in St Louis, Missouri, and a co-author with Fleming on the *C. tepidum* paper, admits to some scepticism. “My sense is that there may well be a few cases, like the ones we know about already, where these effects are important,” he says, “but that many, if not most, biological systems will not utilize quantum effects like these.” But Scholes believes that there are grounds for optimism, given a suitably broad definition of quantum biology. “I do think there are other examples in biology where an understanding at the quantum-mechanical level will help us to appreciate more deeply how the process works,” he says.

## THE BIRD’S-EYE COMPASS

One long-standing biological puzzle that might be explained by exotic quantum effects is how some birds are able to navigate by sensing Earth’s magnetic field.

The avian magnetic sensor is known to be activated by light striking the bird’s retina. Researchers’ current best guess at a mechanism is that the energy deposited by each incoming photon creates a pair of free radicals<sup>8</sup> — highly reactive molecules, each with an unpaired electron. Each of these unpaired electrons has an intrinsic angular momentum, or spin, that can be reoriented by a magnetic field. As the radicals separate, the unpaired electron on one is primarily influenced by the magnetism of a nearby atomic nucleus, whereas the unpaired electron on the other is further away from the nucleus, and feels only Earth’s magnetic field. The difference in the fields shifts the radical pair between two quantum states with differing chemical reactivity.

“One version of the idea would be that some chemical is synthesized” in the bird’s retinal cells when the system is in one state, but not when it’s in the other, says Simon Benjamin, a physicist at the University of Oxford, UK. “Its concentration reflects Earth’s field orientation.” The feasibility of this idea was demonstrated in 2008 in an artificial photochemical reaction, in which magnetic fields affected the lifetime of a radical pair<sup>9</sup>.

Benjamin and his co-workers have proposed that the two unpaired electrons, being created by the absorption of a single photon, exist in

## ➔ NATURE.COM

To learn more about quantum computing, visit: [go.nature.com/u55j5c](http://go.nature.com/u55j5c)



a state of quantum entanglement: a form of coherence in which the orientation of one spin remains correlated with that of the other, no matter how far apart the radicals move. Entanglement is usually quite delicate at ambient temperatures, but the researchers calculate that it is maintained in the avian compass for at least tens of microseconds — much longer than is currently possible in any artificial molecular system<sup>10</sup>.

This quantum-assisted magnetic sensing could be widespread. Not only birds, but also some insects and even plants show physiological responses to magnetic fields — for example, the growth-inhibiting influence of blue light on the flowering plant *Arabidopsis thaliana* is moderated by magnetic fields in a way that may also use the radical-pair mechanism<sup>11</sup>. But for clinching proof that it works this way, says Benjamin, “we need to understand the basic molecules involved, and then study them in the lab”.

### SELECTED BENEFITS

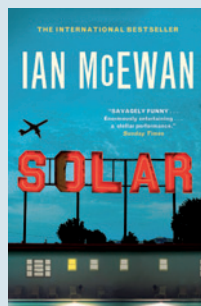
Quantum coherence in photosynthesis seems to be beneficial to the organisms using it. But did their ability to exploit quantum effects evolve through natural selection? Or is quantum coherence just an accidental side effect of the way certain molecules are structured? “There is a lot of speculation about the evolutionary question, and a lot of misunderstanding,” says Scholes, who is far from sure about the answer. “We cannot tell if this effect in photosynthesis is selected for, nor if there is the option not to use coherence to move the electronic energy. There are no data available at all even to address the question.”

He points out that it isn't obvious why selection would favour coherence. “Almost all photosynthetic organisms spend most of the day trying to moderate light-harvesting. It is rare to be light-limited. So why would there be evolutionary pressure to tweak light-harvesting efficiency?”

## A novel idea



### Quantum fact meets fiction



“How your average leaf transfers energy from one molecular system to another is nothing short of a miracle ... Quantum coherence is key to the efficiency, you see, with the system sampling all the energy pathways at once. And the way nanotechnology is heading, we could copy this with the right materials.”

These words are lifted from the pages of Ian McEwan's novel *Solar* (Jonathan Cape, 2010), which describes the tragicomic exploits of physicist Michael

Beard, a Nobel laureate and philanderer, as he misappropriates an idea for a solar-driven method to split water into its elements.

“I wanted to give him a technology still on the lab bench,” says McEwan, who has previously scattered science through his books *Enduring Love* (1997) and *Saturday* (2005). He came across research into quantum photosynthesis by Graham Fleming, a chemist at the University of California, Berkeley, and decided that it was just what he needed. He fit the idea in with Beard's supposed work in quantum physics with the help of Graeme Mitchison, a physicist at the University of Cambridge, UK, who reverse-engineered the Nobel citation for Beard that appears in *Solar's* appendix, and reads, “Beard's theory revealed that the events that take place when radiation interacts with matter propagate coherently over a large scale compared to the size of atoms.” **P.B.**

Fleming agrees: he suspects that quantum coherence is not adaptive, but is simply “a by-product of the dense packing of chromophores required to optimize solar absorption”. Scholes hopes to investigate the issue by comparing antenna proteins isolated from species of cryptophyte algae that evolved at different times.

But even if quantum coherence in biological systems is a chance effect, adds Fleming, its consequences are extraordinary, making systems insensitive to disorder in the distribution of energy. What is more, he says, it “enables ‘rectifier-like’ one-way energy transfer, produces the fastest [energy-transfer]

**“This might just give us a few clues in the quest to create quantum technology.”**

rate, is temperature-insensitive and probably a few other things I haven't thought of”.

These effects, in turn, suggest practical uses. Perhaps most obviously, says Scholes, a better understanding of how biological systems achieve quantum coherence

in ambient conditions will “change the way we think about design of light-harvesting structures”. This could allow scientists to build technology such as solar cells with improved energy-conversion efficiencies. Seth Lloyd considers this “a reasonable expectation”, and is particularly hopeful that his discovery of the positive role of environmental noise will be useful for engineering photonic systems using materials such as quantum dots (nanoscale crystals) or highly branched polymers studded with light-absorbing chemical groups, which can serve as artificial antenna arrays.

Another area of potential application is in quantum computing. The long-standing goal of the physicists and engineers working in this area is to manipulate data encoded in quantum bits (qubits) of information, such as the spin-up and spin-down states of an electron or of an atomic nucleus. Qubits can exist in both states at once, thus permitting the simultaneous exploration of all possible answers to the computation that they encode. In principle, this would give quantum computers the power to find the best solution far more quickly than today's computers can — but only if the qubits can maintain their coherence, without the noise of the surrounding environment, such as the jostling of neighbouring atoms, destroying the synchrony of the waves.

But biology has somehow solved that challenge: in effect, quantum coherence allows a photosystem to perform a ‘best-path’ quantum computation. Benjamin, whose main interest is in designing materials systems for quantum computation and information technology, sees the ambient-temperature avian compass as a potential guide. “If we can figure out how the bird's compass protects itself from decoherence, this might just give us a few clues in the quest to create quantum technologies,” he says. Learning from nature is an idea as old as mythology — but until now, no one has imagined that the natural world has anything to teach us about the quantum world. ■

**Philip Ball** is a writer based in London.

- Engel, G. S. *et al.* *Nature* **446**, 782–786 (2007).
- Lee, H., Cheng, Y.-C. & Fleming, G. R. *Science* **316**, 1462–1465 (2007).
- Panitchayangkoon, G. *et al.* *Proc. Natl Acad. Sci. USA* **107**, 12766–12770 (2010).
- Collini, E. *et al.* *Nature* **463**, 644–647 (2010).
- Mohseni, M., Rebentrost, P., Lloyd, S. & Aspuru-Guzik, A. *J. Chem. Phys.* **129**, 174106 (2008).
- Ball, P. *Nature* **431**, 396–397 (2004).
- Turin, L. *Chem. Senses* **21**, 773–791 (1996).
- Ritz, T., Thalau, P., Phillips, J. B., Wiltschko, R. & Wiltschko, W. *Nature* **429**, 177–180 (2004).
- Maeda, K. *et al.* *Nature* **453**, 387–390 (2008).
- Gauger, E. M., Rieper, E., Morton, J. J. L., Benjamin, S. C. & Vedral, V. *Phys. Rev. Lett.* **106**, 040503 (2011).
- Ahmad, M., Galland, P., Ritz, T., Wiltschko, R. & Wiltschko, W. *Planta* **225**, 615–624 (2007).

# COMMENT

**MEDICINE** Patients' and clinicians' top ten priorities for schizophrenia **p.277**

**BIOLOGY** Zimmer's essay collection on the versatility of viruses **p.279**

**RECORDS** On the beauty and import of field notes **p.281**

**MUSIC** The rap guide to evolution takes to the New York stage **p.282**



ILLUSTRATIONS BY DANIEL MACKIE



## Buried by bad decisions

Our brains are hard-wired to make poor choices about harm prevention in today's world. But we can fight it, says **Daniel Gilbert**.

**T**he London Association for the Prevention of Premature Burial was founded in 1896 to prevent “premature burial generally, and especially amongst the members”<sup>1</sup>. Because nineteenth-century physicians couldn't always distinguish the nearly dead from the really most sincerely dead, premature burial was a problem. But not a big problem. The odds of being buried alive in 1896 were, like the odds of being buried alive today, very close to zero. Nonetheless, the good citizens of England formed action committees, wrote editorials and promoted legislation that ultimately led to expensive safeguards against “the horrible doom of being buried alive”<sup>1</sup>. Most of those

safeguards — such as the costly requirement that bodies spend time in ‘attractive waiting mortuaries’ before being buried — are still with us today. The frequency with which modern cadavers use this waiting period to demonstrate that they've been misdiagnosed is approximately never.

Premature burial isn't a big problem, but the way we deal with big problems is. When an aeroplane's fuselage rips open mid-flight, or an offshore oil rig explodes, or a nuclear power plant is crippled by a tsunami, we immediately ask what could have been done

differently, blame those who didn't do it, then allocate funds and pass legislation to make sure it gets done that way the next time. At first blush, this seems sensible. After all, no one is in favour of aviation accidents, reactor meltdowns or oil spills; so when these things happen, why not do everything we can to make sure they don't happen again?

The answer is that because resources are finite, every sensible thing we do is another sensible thing we don't. Alas, research shows that when human beings make decisions, they tend to focus on what they are getting and forget about what we are forgoing. For example, people are more likely to buy an item when they are asked to choose ▶

► **NATURE.COM**  
Can decision-making be taught?  
[go.nature.com/ykpugo](http://go.nature.com/ykpugo)



► between buying and not buying it than when they are asked to choose between buying the item and keeping their money “for other purchases”. Although “not buying” and “keeping one’s money” are the same thing, the latter phrase reminds people of something they know but typically fail to consider: buying one thing means not buying another. So should we do everything in our power to stop global warming? To make sure terrorists don’t board aeroplanes? To keep *Escherichia coli* out of the food supply? These seem like simple questions with easy answers only because they describe what we will do without also describing what we won’t. When both are made explicit — should we keep hamburgers safe or aeroplanes safe? — these simple questions become vexing. Harm prevention often seems like a moral imperative, but because every yes entails a no, it is actually a practical choice.

How are we to make that choice? In the seventeenth century, Blaise Pascal and Pierre de Fermat derived the optimal strategy for betting on games of chance, and in the process demonstrated that wise choices about harm prevention are always the product of two estimates: an estimate of odds (how likely is the harmful event?) and an estimate of consequences (how much harm will it cause?). If we know which harm is most likely and which harm is most severe, then we know which harm to prevent. We should spend less to prevent a natural disaster that will probably leave 3,000 people homeless than a communicable disease that will certainly leave 3 million people dead, and this is perfectly obvious to everyone.

Except when it isn’t.

### ANCIENT MINDS

The reason it took a pair of mathematical geniuses to develop a formula for rational choice is that human beings often don’t make choices that way. When left to our own devices, we will pay more to eliminate a small risk of illness than to reduce a large one<sup>2</sup>, and more to insure ourselves against a scary way of dying than against every way of dying<sup>3</sup>. We will save all the members of a five-person group before we will save six members of a ten-person group<sup>4</sup>, and we will save lives by pushing a trolley into a person but not a person into a trolley<sup>5</sup>. Our brains were optimized for finding food and mates on the African savannah and not for estimating the likelihood of a core breach or the impact of overfishing. Nature has installed in each of us a threat-detection system that is exquisitely sensitive to the kinds of threats our ancestors faced — a slithering snake, a romantic rival, a band of men waving sticks — but that is remarkably insensitive to the odds and consequences of the threats we face today.

For example, our brains devote a great deal of time and real estate to processing



information about other people — about what they think, know, want and intend. Because we specialize in understanding other minds, we are hypersensitive to the harms those minds produce. When people play economic games, for instance, they tend to reject unfair offers from their opponents — but they are much more likely to do so when their opponent is a person than when their opponent is a computer<sup>5</sup>. When people receive electric shocks, they describe them as considerably more painful when they are intentionally administered by a human agent<sup>6</sup>. It is bad to be harmed, but it is worse to be victimized. And so we worry more about shoe-bombers

**“We will change our lives to save a child but not our light bulbs to save them all.”**

than influenza, despite the fact that one kills roughly 400,000 people per year and the other kills roughly none. We worry more about our children being kidnapped by strangers than about becoming obese, despite the fact

that abduction is rare and diabetes is not. Terrorists and child-molesters are agents, viruses and French fries are objects, and agents threaten us in a way that objects never can.

We are especially concerned when the threats those human agents produce are to our dignity, values and honour. Moral rules bind communities together, enable trust and the division of labour and cause people to behave honestly when no one is watching. Because these rules have such a crucial role in the formation and functioning of human social groups, we are obsessed with their violation, which is why *US Weekly* outsells *The New Yorker*. Unfortunately, when a tribe grows to nearly 7 billion people, threats to its sense of decency are not the most serious threats it faces. Climate change is caused by

the burning of fossil fuels, not flags. Because a decision to prevent one kind of harm is always a decision not to prevent another, the irresistible lure of moral violations can distract us from more crucial concerns.

### MORALS TO DIE FOR?

Our obsession with morality can also discourage us from embracing practical solutions to pressing problems. The taboo against selling our bodies means that people who have money and need a kidney must die so that people who need money and have a spare kidney can starve. Economic models suggest that drug abuse would decline if drugs were taxed rather than banned<sup>7</sup>, but many people have zero tolerance for policies that permit immoral behaviour even if they drastically reduce its frequency. Licensing prostitutes, trading pollution credits and paying students to stay in school may or may not reduce harm, but many would oppose these ideas even if they were proved effective. It is apparently better for people to suffer and die than to get the wrong message.

Our species’ sociality has always been its greatest advantage, but it may also be its undoing. Because we see the world through a lens of friends and enemies, heroes and villains, alliances and betrayals, virtue and vice, credit and blame, we are riveted by the dramas that matter least and apathetic to the dangers that matter most. We will change our lives to save a child but not our light bulbs to save them all.

What are we to do about the mismatch between the way we think and the problems we should be thinking about? One solution is to frame problems in ways that appeal to our nature. For example, when threats are described as moral violations, apathy often turns to action. Texas highways were awash in litter until 1986, when the state adopted

a slogan — ‘Don’t mess with Texas’ — that made littering an insult to the honour of every proud Texan, at which point littering decreased by 72% (ref. 8). Hotels wasted significant amounts of energy washing barely-used towels until 2008, when researchers placed signs in hotel rooms that either asked guests to “help save the environment by reusing your towels” or told guests that “75% of the guests who stayed in this room participated in our new resource savings program by using their towels more than once”<sup>9</sup>. The second sign suggested that laundering a barely-used towel was a violation of a moral rule that most people obeyed, and that sign increased towel reuse by 33%. Psychologists and economists have found dozens of ways to make problems easier to think about and harder to ignore. There is no shortage of solutions, just of the will to implement them.

The other way to deal with the mismatch between the threats we face and the way we think is to change the way we think. People are capable of thinking rationally about odds and consequences, and it isn’t hard to teach them. Research shows that a simple five-minute lesson dramatically improves people’s decision-making in new domains a month later<sup>10</sup>, and yet that is five minutes more than most people ever get. We teach high-school students how to read Chaucer and do trigonometry, but not how to think rationally about the problems that could extinguish their species.

Psychologists have made remarkable progress in understanding how decision-making goes wrong and how it can be set right, and although their research generates bestselling books and garners Nobel Prizes, funding agencies typically give it low priority. Our communal fate rests on decisions that could easily be improved, if only we would decide to do so. It is our way of thinking, and not the undertaker, that threatens to bury us prematurely. ■

**Daniel Gilbert** is in the Department of Psychology, Harvard University, Cambridge, Massachusetts 02138, USA.  
e-mail: gilbert@wjh.harvard.edu

1. *J. Am. Med. Assoc.* **XXXII**, 947 (1899).
2. Viscusi, W. K., Magat, W. A. & Huber, J. *Rand J. Econ.* **18**, 465–479 (1987).
3. Johnson, E. J., Hershey, J., Meszaros, J. & Kunreuther, H. *J. Risk Uncertain.* **7**, 35–51 (1993).
4. Jenni, K. & Loewenstein, G. *J. Risk Uncertain.* **14**, 235–257 (1997).
5. Sanfey, A. G., Rilling, J. K., Aronson, J. A., Nystrom, L. E. & Cohen, J. D. *Science* **300**, 1755–1758 (2003).
6. Gray, K. & Wegner, D. M. *Psychol. Sci.* **19**, 1260–1262 (2008).
7. Becker, G. S., Murphy, K. M. & Grossman, M. J. *Polit. Econ.* **114**, 38–60 (2006).
8. McClure, T. & Spence, R. *Don’t Mess with Texas: The Story Behind the Legend* (Idea City Press, 2006).
9. Goldstein, N. J., Cialdini, R. B. & Griskevicius, V. *J. Consum. Res.* **35**, 472–482 (2008).
10. Larrick, R. P., Morgan, J. N. & Nisbett, R. E. *Psychol. Sci.* **1**, 362–370 (1990).

# Democratizing clinical research

**Keith Lloyd and Jo White** commend a way for patients, clinicians and scientists to set priorities jointly.

**R**esearch priorities are rarely set democratically. Whereas clinical science is largely about establishing which treatments work best for whom, sadly, the views of those with most to gain or lose — patients — are generally ignored. Academics, industry and other big players with vital roles in developing treatments tend to set the agenda. But their priorities differ from those of patients and clinicians. For example, the outcomes measured in a trial of a drug may not be those of interest to the people who will actually take it.

The inclusion of patient demands is not a panacea. It can divert scarce research resources and delay important treatments<sup>1</sup>. One solution is to try to harmonize the perspectives of patient and clinician. This is what the James Lind Alliance (JLA) Priority Setting Partnerships in Oxford, UK, attempt, perhaps uniquely. Established in 2004 and funded by the UK Medical Research Council and National Institute for Health Research (NIHR), the JLA

brings together patients, carers and clinicians to identify and rank questions about the effects of treatments for a given disease. Clinicians and academics — who may never meet patients — find long-held beliefs challenged and sometimes overturned.

The JLA process has recently been applied to schizophrenia — a mental illness affecting about one person in a hundred worldwide. We were involved in this exercise as clinical academics. This, plus our experience as recipients of grants and from within funding bodies, convinces us that money rarely goes to the studies that those with mental illness would choose. We therefore urge funders to adopt this list of top priorities for schizophrenia (see ‘Top ten treatment uncertainties’), and entreat other countries and organizations to use the technique involved in compiling it to steer other clinical research.

Between 2007 and 2009, we and other collaborators from the JLA Partnership collated 489 potential uncertainties about

## SCHIZOPHRENIA RESEARCH PRIORITIES

### Top ten treatment uncertainties

1. What is the best way to treat people with schizophrenia that is unresponsive to treatment?
2. What training is needed to recognize the early signs of recurrence?
3. Should there be compulsory community outpatient treatment for people with severe mental disorders?
4. How can sexual dysfunction due to antipsychotic-drug therapy be managed?
5. What are the benefits of supported employment for people with schizophrenia in terms of quality of life, self esteem, long-term employment prospects and illness outcomes?
6. Do the adverse effects of antipsychotic drugs outweigh the benefits?
7. What are the benefits of hospital treatment compared with home care for psychotic episodes?
8. What are the clinical benefits and cost-effectiveness of monitoring the physical health of people with schizophrenia?
9. What are the clinical, social and economic outcomes — including quality of life and the methods and effects of risk monitoring — of treatment by acute day hospitals, assertive outreach teams, in-patient units, and crisis resolution and home treatment teams?
10. What interventions could reduce weight gain in schizophrenia?

Some treatment uncertainties have been reformulated here as questions.

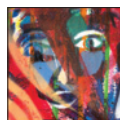


the treatment of schizophrenia. These came from clinicians, patients and their carers through web- and paper-based questionnaires. We also pulled them from the UK Database of Uncertainties about the Effects of Treatments, which contains instances in which “no up to date systematic reviews exist, or up-to-date systematic reviews show that uncertainty continues”.

These questions were de-duplicated to produce a longlist of 237 issues. Eleven schizophrenia partners — carers, clinicians, patients, funders and voluntary-sector organizations — each ranked their top ten uncertainties. These partners responded either as individuals, or on behalf of an organization, having consulted colleagues and members.

The partnership collated the rankings, recording separate running totals for patient, carer and clinician submissions. This enabled a steering group — a subset of the partners — to examine each individual ranking, as well as the combined ranking, to produce a pooled list of 26 treatment uncertainties.

Finally, this list was discussed at an exhilarating workshop of clinicians, carers, patients, funders and voluntary-sector organizations in January. The JLA facilitated the meeting using a structured variation of small-group discussion called ‘nominal group technique’ (see [go.nature.com/xswwtc](http://go.nature.com/xswwtc)) to



## SCHIZOPHRENIA

Search for origins and treatments  
[nature.com/schizophrenia](http://nature.com/schizophrenia)

reach moderated consensus on a top ten.

The process prevented one person dominating the discussion and encouraged all group members to participate. The format was rigorous, but flexible enough to allow people to revise their opinions, raise concerns and to reach consensus about any imbalance perceived to have emerged from the interim stages.

Although the purpose of the JLA process<sup>2</sup> is to enable patients and those who treat them to have a say in what gets studied, it can also change clinical practice. For example, sexual dysfunction caused by antipsychotic medication emerged as a key patient priority. This is typically a low priority for clinicians prescribing medication and for companies assessing drug effectiveness.

The week after the JLA workshop, a patient came to see one of us (K.L.) in a clinic, and wanted a change of antipsychotic medication because of sexual dysfunction. Without the experience of the JLA process, it is unlikely that this issue would have been afforded as much weight as it was.

The final top ten for schizophrenia is

noteworthy for its divergence from the agenda of the drug industry, and begs many questions. Perhaps most pressing: is it ethical to conduct research, which may include testing new treatments, without considering which outcomes matter most to those who will receive the treatment? And is it, in the long run, to drug companies' benefit to do so? Such questions are particularly pertinent in conditions such as schizophrenia, in which the balance of power between researcher, clinician and patient is so uneven.

What next? The team will repeat the exercise for depression this year and next. Meanwhile, the JLA is encouraging funders and researchers to act on the top ten rather than to continue with agendas devoid of clinician and patient input. For example, the NIHR is now exploring commissioning research on weight gain and sexual dysfunction in schizophrenia. Assumptions that “researcher knows best” have had their time. ■

**Keith Lloyd and Jo White** are at the College of Medicine, Swansea University, Swansea SA2 8PP, UK.

*e-mail:* [k.r.lloyd@swansea.ac.uk](mailto:k.r.lloyd@swansea.ac.uk)

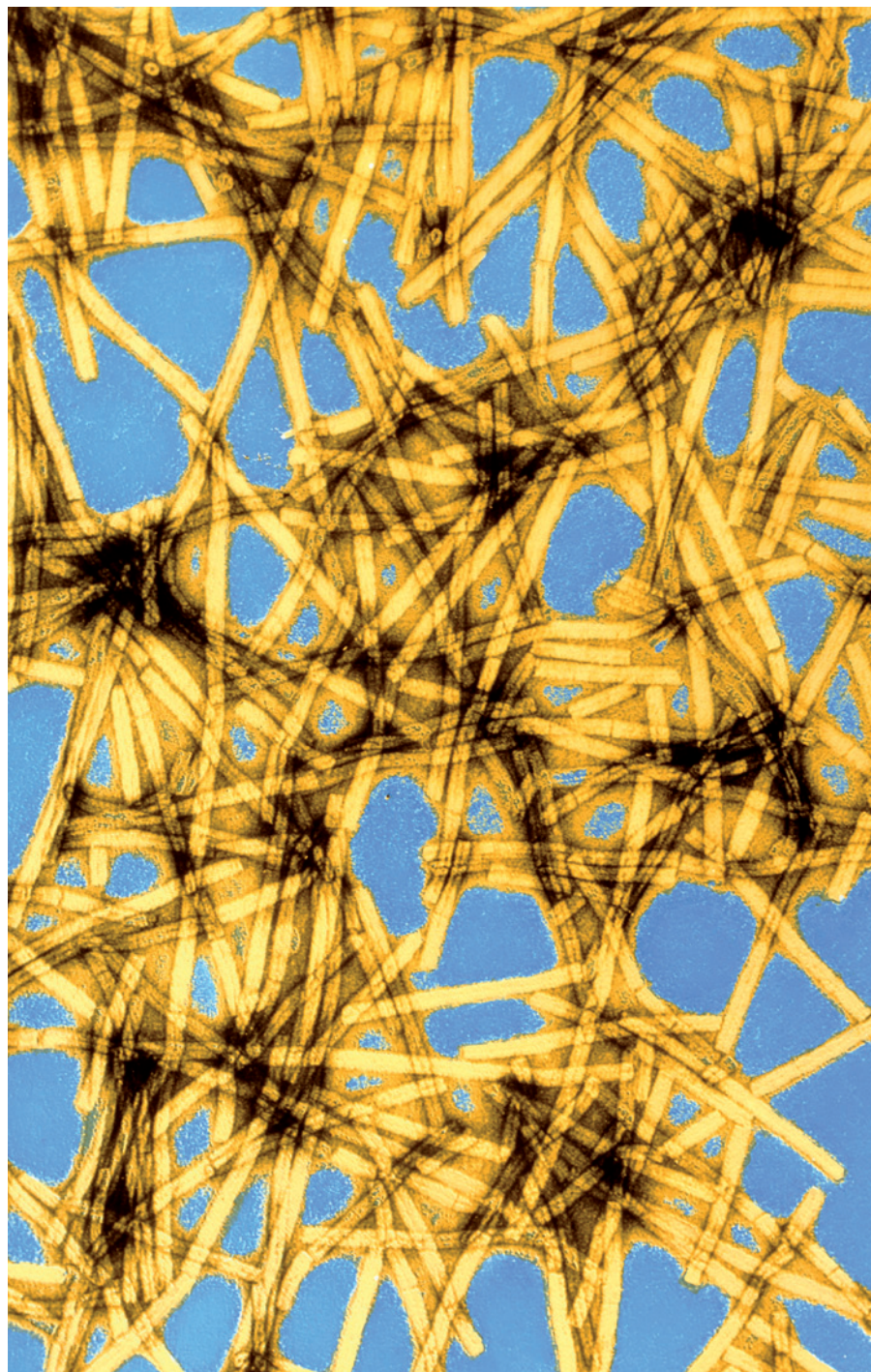
1. Chafe, R. *et al.* *Nature* **472**, 410–411 (2011).
2. Cowan, K. & Oliver, S. *The James Lind Alliance Guidebook* (2011); available at [go.nature.com/u2cqcy](http://go.nature.com/u2cqcy)

## VIROLOGY

# Potent tiny packages

Carl Zimmer's primer on viruses entertains, but reveals little about their basic traits, says **Robin Weiss**.

J. BURGESS/SPL



Rods of tobacco mosaic virus were first seen in 1939 using an electron microscope.

Viruses propagate in every kind of living organism and, despite being so tiny, amount to around 5% of the world's biomass. Viruses that infect algal blooms, for example, induce the calcification of their hosts and are thus responsible for the white cliffs of Dover in Britain, and other carbon sinks. Acclaimed science writer Carl Zimmer celebrates the versatility of these agents in a dozen entertaining essays in his latest book, *A Planet of Viruses*, which accompanies the World of Viruses educational website ([www.worldofviruses.unl.edu](http://www.worldofviruses.unl.edu)).

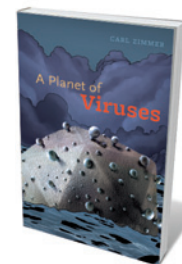
Each essay deals with a different virus, ranging from those in bacteria and plants to scourges such as smallpox, severe acute respiratory syndrome (SARS), influenza and West Nile virus. There is food for thought here for all, even a seasoned virologist like me. Zimmer's writing grabs one's interest, but is marred by a lack of attention to detail. It would be a more attractive little volume if it were half as long again, because many of the essays end just when they become interesting.

Zimmer begins with the observation in 1898 by the Dutch microbiologist Martinus Beijerinck that the tobacco mosaic disease of plants was caused by "a contagious living fluid". Dmitry Josifovich Ivanovsky is ignored in the account — even though he had also isolated and propagated tobacco mosaic virus from filtered plant sap six years earlier — because he thought it must be a bacterial disease. The property of transmission through a filter too fine for bacteria to pass through became the defining feature of a virus, thanks to Beijerinck but also to Friedrich Loeffler and Paul Frosch, who reported the filterable nature of foot-and-mouth disease in the same year.

Zimmer omits the marvellous subsequent discoveries using tobacco mosaic virus: its crystallization from the 'living fluid' by Wendell Stanley in 1935; its composition as protein wrapped around RNA by Norman Pirie and Frederick Bawden in 1936; the first visualization of virus particles through the electron microscope by Gustav Kausche, Edgar Pfannkuch and Helmut Ruska in 1939; and the first demonstration, by Heinz Fraenkel-Conrat in 1955, that RNA alone can reconstitute infection.

One of the most fascinating tales of twentieth-century science is how viruses opened up molecular biology, but you won't find it here.

This year marks the centenary of Peyton Rous's demonstration that a cancer in chickens can be transmitted by a filterable agent. Zimmer's essay 'Rabbits with



**A Planet of Viruses**  
CARL ZIMMER  
University of Chicago  
Press: 2011. 109 pp.  
£13, \$20



► Horns' recalls Rous's pioneering work on tumour viruses, before introducing Richard Shope's discovery of the rabbit papilloma viruses, to which the essay title refers. Harald zur Hausen and colleagues discovered human cervical papilloma viruses in 1983; the vaccines that target them to protect women against cervical cancer were licensed in 2006. Thus, 100 years of tumour virology paid off. Rous may have had to wait longer than any other Nobel laureate to win his prize in 1966, whereas zur Hausen's 'incubation period' was a mere 25 years.

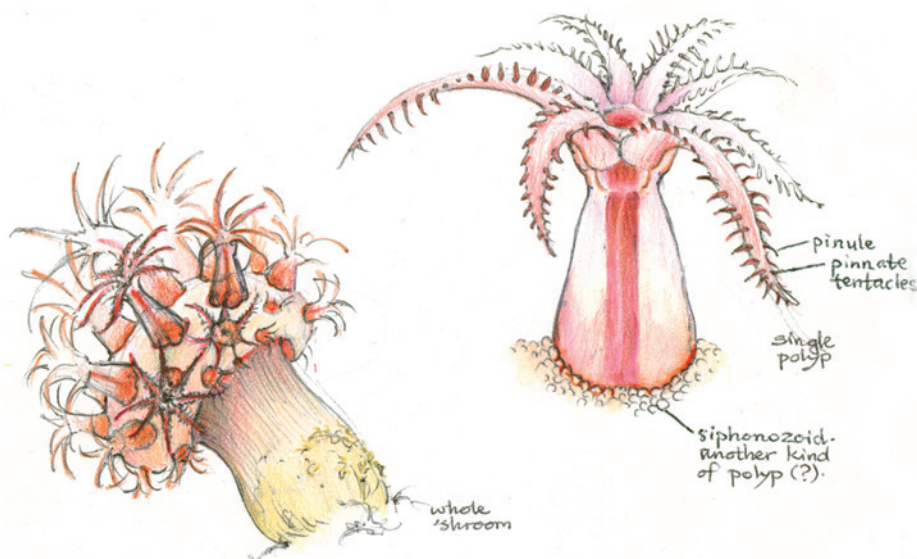
Owing to Zimmer's puzzling reluctance to delve into molecular virology, we have to wait until the end of the last and best essay, on the giant mimiviruses — discovered only in 1992 — to learn that some viruses have RNA genomes instead of DNA. No other replicating systems carry their genes in the form of RNA, as do polio, measles, influenza and most plant viruses. Some viruses have double-stranded RNA, whereas others are single-stranded; some viruses carry a single RNA or DNA molecule, and others have segmented genomes like the different chromosomes of higher organisms. Neither of the two chapters on retroviruses mentions reverse transcription — by which the RNA genome is turned into DNA before inserting itself into host DNA — even though the most potent anti-HIV drugs are designed to block this process.

Perhaps Zimmer thinks such facts are too difficult for his readership, but I view avoiding them as dumbing down. Which viruses evolved from bacteria, and which are more likely to have emerged as sets of genes that escaped from their hosts? Are some viruses relics from an RNA-based world, or are they relatively modern parasites derived from other living systems? Zimmer eventually raises the last question, but to my mind, the fascination of viruses is their enormous molecular and evolutionary diversity as much as their pervasiveness in the environment.

Concern for accuracy seems to have suffered as Zimmer becomes an ever more prolific writer. Some virologists' names are wrong, for instance, as are some other simple facts.

In a foreword to Peter Medawar's 1996 collection of essays, Stephen Jay Gould called this literary form "a weapon of wit and instruction". It would be difficult for any science writer to match Medawar or Gould; nevertheless, Zimmer's contributions fit this definition well. ■

**Robin Weiss** is professor of viral oncology at University College London, UK.  
e-mail: r.weiss@ucl.ac.uk



J. KELLER

A mushroom coral and one of its polyps drawn in the field by science illustrator Jenny Keller.

#### TECHNIQUES

## Records in the field

Good notebook skills are vital for documenting observations of the natural world, finds **Sandra Knapp**.

**F**ield biology: the very words conjure up romance, danger, excitement. There is a thrill to fieldwork that makes lab-based scientists ask "How was your holiday?" when one returns from a stint outside. Many books have been written about the explorers of the past, transcribing their logs and journals, or fictionalizing their adventures. This volume is refreshingly different. Biologist Michael Canfield has compiled a set of essays not on researchers' travels, but on how they capture their experiences in their notes.

*Field Notes on Science & Nature* is an eclectic collection that crosses many disciplines, from geology, botany and zoology to art and anthropology. The variety of styles and records described are fascinating — field notes are very personal. Some of the contributors take notes entirely electronically, others in red pen in cheap notebooks. Others use pictures more than words.

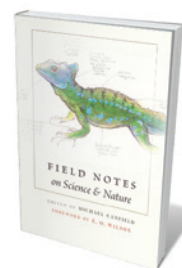
Few of us have the artistic skills of Jonathan Kingdon or Jenny Keller, scientist-illustrators whose drawings alone make this book worth buying. But even the sketchiest sketch can call to mind a place or organism in a way no words can. I remember the field books of a friend with whom I worked in the tropical forests of Central

America: an incredible mixture of description, sketches and taped-in leaves. Today, his

seemingly chaotic collections evoke those places far better than my own lists. I learned from this, and started to sketch the plants I collected — flower and leaf shapes, plant forms and outlines appeared in my pages.

The tradition has a long pedigree, encompassing notebook sketches by the great Victorian naturalists. My favourites are those of Henry Walter Bates, Alfred Russel Wallace and Richard Spruce, early evolutionists who mused on the page about why, as well as what and where. Keller's advocacy of standardized colour palettes in her essay harks back to the methods of eighteenth-century illustrator Sydney Parkinson, who accompanied Joseph Banks on Captain James Cook's voyage on HMS *Endeavour*, or the Austrian Bauer brothers, one of whom accompanied Captain Matthew Flinders on HMS *Investigator* a few decades later.

Parkinson drew and painted all of the plants that were collected, but for efficiency only coloured part of each (a practice recommended by Keller). He died on the voyage,



**Field Notes on Science & Nature**  
EDITED BY MICHAEL R. CANFIELD  
Harvard University Press: 2011. 320 pp.  
\$27.95

► **NATURE.COM**  
For more on  
visualization:  
[go.nature.com/fhvacz](http://go.nature.com/fhvacz)

but his work was enough to enable the publication of the entire collection two centuries later as a series of coloured plates. Ferdinand Bauer's sketches of plants and animals of Australia were intricately labelled with numbers indicating colours; it was only in the twentieth century that the key to the colours was discovered, deep in the collections in Madrid. His brother, Franz, used the same key in botanical paintings he made at Kew, near London.

Accuracy and speed of capture of the image are just as important now. But digital photography has not obviated the need for field sketches. As many contributors point out, a sketch can be labelled on the spot and does not require printers, cameras and other electronic hardware to be carried to remote places.

Whether notes are telegraphic or detailed, a key to abbreviations is a must. Making field notes directly on the computer can solve the transcription problem, as one only has to enter information once and typed text is easy to read, say entomologist Piotr Naskrecki and plant biologist Jim Reveal. But, Reveal adds, computerized notes lack the personality so apparent in handwritten accounts.

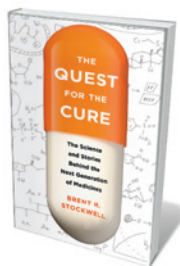
Illustrated field notes can provide the basis for public conversations on science. For example, anthropologist Karen Kramer's sketch maps of Mayan villages aided her research into how the villages functioned because local people were happy to talk about her interpretations of their space. And ornithologist Kenn Kaufman describes the species lists made through the eBird project, which records birders' observations via a website. This crowd-sourcing method of taking field notes is an extension of the 'bioblitz' concept, in which members of the public list all the species they encounter over a short period.

It is disturbing to observe, as ecologist Erick Greene does in his essay on best practice, that today's generation of field biologists do not keep notes as diligently as their laboratory-based counterparts. Lab books are retained as permanent records (sometimes drawn upon in cases of scientific misconduct), whereas field notebooks are rarely archived. Yet they record observations that might seem trivial at the time, but on reflection become the basis for new insight. As ecologist Bernd Heinrich rightly says, notes from the field often represent a search for problems, not solutions. Who knows whose field notebooks now contain observations that will change the world?

I will alter my own note-taking after reading this set of essays. All scientists, whether based in the field or the lab, could benefit from the advice given here so eloquently. ■

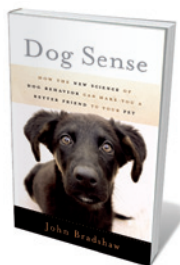
**Sandra Knapp** is a botanist at the Natural History Museum, London SW7 5BD, UK.  
e-mail: s.knapp@nhm.ac.uk

## Books in brief



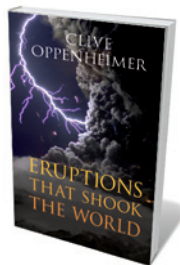
### The Quest for the Cure: The Science and Stories Behind the Next Generation of Medicines

*Brent R. Stockwell* COLUMBIA UNIVERSITY PRESS 284 pp. \$27.95 (2011)  
In the past 50 years, we have developed drugs to cure many major diseases. But treatments for some serious conditions, such as cancer and Alzheimer's, still elude us. Chemical biologist Brent Stockwell describes the history of drug design, from the invention of mustard gas and early anti-cancer agents to the decoding of the human genome. Countering the pessimists who fear that the end is nigh for significant breakthroughs, he argues that emerging technologies for drug testing and molecular modelling will open up new avenues.



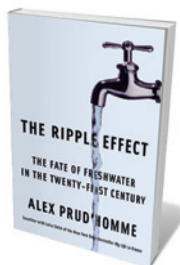
### Dog Sense: How the New Science of Dog Behavior Can Make You A Better Friend to Your Pet

*John Bradshaw* BASIC BOOKS 352 pp. \$25.99 (2011)  
Although dogs are loved by many, their lot is not always a happy one. Originally bred as rural working animals, most dogs now live in cities where they are expected to be more obedient than any child. The perpetuation of pedigrees also mars the health of some breeds. Anthrozoologist John Bradshaw summarizes what science can teach us about man's best friend. Arguing that modern dogs should not be considered domesticated wolves, he asks how we can best breed these social animals to be companions and family pets.



### Eruptions that Shook the World

*Clive Oppenheimer* CAMBRIDGE UNIVERSITY PRESS 408 pp. £18.99 (2011)  
Closures of international airspace after the recent Icelandic eruptions served as a reminder that volcanoes can be disruptive. But volcanic outbursts have also shaped our history, from aiding the demise of the dinosaurs to altering climate. Ash ejected into the atmosphere may even have led to the meagre harvest that triggered the French Revolution. Volcanologist Clive Oppenheimer relates in rigorous detail the consequences of eruptions over the past quarter of a billion years, and argues that lessons can be learned for future risk management of catastrophes.



### The Ripple Effect: The Fate of Freshwater in the Twenty-First Century

*Alex Prud'homme* SCRIBNER 448 pp. \$27 (2011)  
Flooding and drought are both on the rise. Journalist Alex Prud'homme digs into the reasons why, citing centuries of neglect of water infrastructure and a careless attitude to issues of water quality and use, ownership and waste. Focusing on issues that threaten clean and abundant water in the United States, he travels across the country to speak to people at the centre of the drama, including salmon fishermen and copper miners in Alaska and scientists investigating intersex fish in Chesapeake Bay.



### The Fallacy of Fine-Tuning: Why the Universe is Not Designed For Us

*Victor J. Stenger* PROMETHEUS 345 pp. £24.95 (2011)  
The Universe seems to be fine-tuned, with precisely set parameters that allow life to exist as a rare event. This idea has been used by some to argue that humans have a central place in the cosmos, and even as evidence for the existence of God. Physicist Victor Stenger rails against this 'fallacy' by dismantling such assumptions one by one. The laws of physics and cosmology constrain some key numbers, he says, and others are not as fine-tuned or as improbable as proponents of the idea suggest.





A. MELTON

## Q&A Baba Brinkman

# The adaptive lyricist

*Baba Brinkman is a Canadian rap artist whose award-winning show The Rap Guide to Evolution wowed UK crowds at the Edinburgh Fringe Festival during Charles Darwin's bicentenary year. As the show opens next week for a long summer run off Broadway in New York, Brinkman discusses rhyme, improvisation and scientific certainty.*

### Why rap about science?

Science chose me. I have a master's degree in medieval literature and I had done a hip-hop show based on *The Canterbury Tales*. Mark Pallen, a bacteriologist at the University of Birmingham, UK, heard it and got it into his head that the next one should be a rap of *On the Origin of Species*. I took up the gauntlet. Science rapping is not the reason I got into rap, but I've found myself evolving into that niche.

### Is it a comfortable niche?

Yes, science and rap go well together. Rap, in essence, is about speaking with conviction — and you can usually be certain of things that have been scientifically validated. I've upped my swagger since I've been writing about science. In my songs from five or

six years ago, when I was immersed in the humanities, I was a big equivocator. Not that I'm trying to convey absolute certainty,

### The Rap Guide to Evolution

BABA BRINKMAN.  
DIRECTED BY DODD  
LOOMIS.  
SoHo Playhouse,  
New York.  
26 June–2 October  
2011.

because science is about uncertainty and exploration.

### Do researchers mind you talking about science with certainty?

I haven't had any scientists come after me because of this show. I've covered my bases — for each thing I mention in my raps I can point to the research that made me want to highlight that element. What is 100% me is the stylistic choice, the decision to expose this part of the research over that part. That is based on entertainment value. I don't have to rap about the formation of quartz crystals.

### What's your favourite piece from your new show?

One that goes to the heart of the matter is 'Performance, Feedback, Revision', which each night contains a freestyle improvised rap (see [go.nature.com/ubvkdz](http://go.nature.com/ubvkdz)). The point is that

mutation is similar to artistic improvisation. Without that randomness, it is difficult to create new material that goes in surprising directions — and leads the charge on the evolutionary development of the artist, or the organism. I dramatize the evolving process by having an unscripted rap that the audience can see is happening in real time. I never know what it's going to be.

### Can you give a sample of the lyrics from the non-improvised part of that song?

"Yeah, you're just a phenotype, performing all the genes inside / Living things only seem designed, 'cause you can't see how they've been revised / And the feedback lies in evolution's brutal gaze / Either you have babies who have babies or get booed off stage."

### What are you going to tackle next?

There is interest in a rap guide to climate change, which would be a good challenge. Converting people to looking at how evolution works and accepting it as a reality is an intellectual battle that is worth fighting. Accepting anthropogenic climate change as a reality is an important social and political agenda to lend my weight to. But it looks like there's going to be a Rap Guide to Business first. New York University's Stern School of Business has expressed an interest in commissioning me to summarize its MBA programme. ■

INTERVIEW BY KERRI SMITH

### ➔ NATURE.COM

Listen to Baba Brinkman on the Nature Podcast: [go.nature.com/qeaijm](http://go.nature.com/qeaijm)

# CORRESPONDENCE

## Extinctions: conserve not collate

Fangliang He and Stephen Hubbell correct an overestimation of 160% for species extinction rates resulting from habitat destruction (*Nature* **473**, 368–371; 2011). However, near-term extinction rates predicted by the Millennium Ecosystem Assessment still remain at 400–4,000 times the background rate of species extinction.

Although it may help to refine future predictions, we caution against their recommendation for collating more detailed geographical data as an urgent priority for conservation science.

Knowing where species occur and their risk of extinction is fundamental for deciding where to focus efforts to protect them. But the diminishing returns on the value of biological surveys (H. S. Grantham *et al.* *Conserv. Lett.* **1**, 190–198; 2008) means that more data may not translate into significantly better decisions. Heterogeneity in the costs and likelihood of success of conservation actions can influence investment priorities far more.

Areas designated a priority for species protection, identified using the 'species-area relationship', are not affected by model uncertainty, taxonomic group or the non-random distribution of species (M. C. Evans *et al.* *Divers. Distrib.* **17**, 437–450; 2011).

**Megan Evans, Hugh Possingham, Kerrie Wilson**  
*The University of Queensland, Australia. m.evans1@uq.edu.au*

## Extinctions: consider all species

We question Fangliang He and Stephen Hubbell's claim that species-area relationships overestimate global extinction



(*Nature* **473**, 368–371; 2011). We contend that they do not test their claims against real data on global extinction or threat. We also believe that they address only a small part of the problem.

Imagine destruction that wipes out 95% of habitat overnight — metaphorically speaking. How many species will have disappeared the following morning? He and Hubbell tell us it would be just those living only in the destroyed area, and not in the other 5%. In our view, the more important question is how many species in total, including those in the remnant habitat 'islands' (the 5%), will eventually become extinct (see M. L. Rosenzweig *Species Diversity in Space and Time* Cambridge Univ. Press, 1995.)

Many studies accurately verify extinction predictions based on the relationship between island area and numbers of species, which He and Hubbell dismiss. Scores of separate tests find striking agreement between the number of predicted extinctions from habitat loss and the number of consequent extinctions (or of species facing extinction). This is seen globally and within individual regions, including eastern North America, South America, Africa and southeast Asia (see, for example, S. L. Pimm and R. A. Askins *Proc. Natl Acad.*

*Sci. USA* **92**, 9343–9347; 1995).

Comprehensive analyses can now combine remotely sensed ecosystem changes with information on species extinction risk, distribution, habitats, threats and conservation actions from the International Union for Conservation of Nature Red List. In our opinion, it is these studies — which ask the right questions and verify the answers — that have crucial implications for the world's efforts to conserve biodiversity.

**T. M. Brooks\*** *NatureServe, Virginia, USA.*

*tbrooks@natureserve.org*

*\*On behalf of 7 co-signatories (see go.nature.com/tsnlzs).*

## Making society more resilient

Japan's government would do well to consider how society can adapt to cope with the uncertainty and change caused by sudden disastrous natural events — called resilience thinking — rather than simply trying to overcome and eliminate such changes.

Catastrophic disturbances such as tsunamis, wildfires, flooding and volcanic eruptions can exact a huge human cost. But they may also have a positive impact on ecosystems, particularly those

eroded by human activity. The 2004 Indian Ocean tsunami, for example, restored the beach nesting habitats for several threatened sea-turtle species (D. B. Lindenmayer and C. R. Tambiah *Conserv. Biol.* **19**, 991; 2005).

The ability of ecosystems to absorb natural disturbances and society's ability to resist and recover from them are connected. History shows that socio-ecological systems that are resilient to hazards are less devastated by recurring natural events such as hurricanes (W. N. Adger *et al.* *Science* **309**, 1036–1039; 2005). Ignoring the connection could lead to more unforeseen economic disasters.

**Akira S. Mori** *Yokohama National University, Japan.*  
*akkym@kb3.so-net.ne.jp*

## Population decline is a long way off

Fred Pearce uses strong words to criticize the United Nations' latest projected global population figures (*Nature* **473**, 125; 2011). But the UN's projections of a continuing rise in the population (see go.nature.com/wj3br5) are in line with its previous projections and with those of other major sources, including the US Census Bureau (see go.nature.com/owcela) and the International Institute for Applied Systems Analysis (go.nature.com/cbg34l).

The new UN 'medium variant' projection expects 10.1 billion people by 2100, 3 billion more than now. This is a sobering prospect for those concerned with human and environmental poverty.

In his book *The Coming Population Crash* (Beacon Press, 2010), Pearce predicts a drastic population decline owing to falling fertility. But the birth rate worldwide still exceeds the replacement rate, so the young greatly outnumber the old. The number of young women coming



into reproductive age can be three times the number becoming post-menopausal. So, although women are now having fewer children than they did previously, the number of children remains high. The US Census Bureau projects no decline in the global number of births to 2050.

The result is that the population has risen by a billion people in the past 13 years and the UN's medium variant expects about the same in the next 13 years.

None of the UN scenarios envisages a rise in fertility. If fertility stays at its present level, the UN projects 27 billion people in 2100. Only by assuming a continuing and rapid fall in fertility do projections come down to between 6 and 16 billion.

Globally, there are 2.5 births for each death (see [go.nature.com/ows9ux](http://go.nature.com/ows9ux)). Population stability, let alone a decline, is therefore a long way off. For the foreseeable future, the world is going to be much more crowded than it is now.

**Robert Wyman** *Yale University, Connecticut, USA.*  
[robert.wyman@yale.edu](mailto:robert.wyman@yale.edu)

## Brazilian soya: the argument for

Your scepticism about a market-based approach to conservation in the Amazon is ill-founded (*Nature* **472**, 5–6; 2011). It is based on a misrepresentation of the partnership in Brazil's Santarém region between US agricultural giant Cargill and environmental group The Nature Conservancy.

The aims of the Santarém partnership are explicitly environmental, not social as you claim. It was set up to reduce deforestation by enforcing Brazil's Forest Code (a federal law restricting the amount of deforestation) and the soya bean moratorium (a voluntary agreement by agribusiness not to source soya from land deforested after 2006).

The partnership monitors farmers' land-use practices in Santarém by satellite and by visits on the ground. Its contribution is crucial in the absence of a legal mechanism to enforce the

soya moratorium and, given the limited government resources, the Forest Code.

Soya production in Santarém comprises less than 0.5% of the total production of the Legal Amazon (<http://sidra.ibge.gov.br>), yet this small region receives intense scrutiny from scientists and the media. Despite this, no evidence has emerged that the partnership has failed to deter deforestation. We must therefore consider what the environmental outcome would have been had The Nature Conservancy not intervened.

**Rachael Garrett** *Stanford University, California, USA.*  
[rachaelg@stanford.edu](mailto:rachaelg@stanford.edu)

## Brazilian soya: the argument against

Rachael Garrett's arguments for a market-based approach to Amazon conservation (see above: *Nature* doi:10.1038/474285a; 2011) hinge on the assumption that the expansion of agro-industrial development in Amazonia is inevitable. Using market mechanisms to solve environmental problems is questionable when those problems are themselves caused by market-driven expansion.

It is the relatively small soya-production area of Brazil's Santarém region that makes it an important case study. If voluntary market-based conservation programmes do not work even on a small scale, what are the chances of success for larger-scale programmes such as the Round Table on Responsible Soy (see [go.nature.com/jc6ua1](http://go.nature.com/jc6ua1)), hailed as the way to mitigate problems created by agro-industry?

Conservation organizations must face up to the social consequences of their programmes. The Santarém case shows that exclusively addressing environmental aspects of a complex problem exacerbates socio-political issues. The social unrest there correlates with environmental degradation in the region (C. S. Simmons *et al. Ann. Assoc. Am. Geogr.* **97**, 567–592; 2007).

Amazonian deforestation has

accelerated and extraction of its resources have continued under the market-based conservation paradigm. It is time for a radical rethink of the development model.

**Brenda Baletti** *University of North Carolina, Chapel Hill, USA.*  
[bbaletti@email.unc.edu](mailto:bbaletti@email.unc.edu)

## Peer reviews: some are already public

Several journals are already making anonymized reviewers' reports public for published papers, as Daniel Mietchen proposes (*Nature* **473**, 452; 2011). These include *Atmospheric Chemistry and Physics* (see [go.nature.com/qamrfc](http://go.nature.com/qamrfc)) and *The EMBO Journal* (see *Nature* **468**, 29–31; 2010). But at the European Molecular Biology Organization, we do not see an equitable way to publish referee reports on rejected manuscripts.

Instead, we favour the transfer between journals of rejected manuscripts, along with full referee reports that could be made public after acceptance of the paper. An extension of this might be to release referee names after several years, or to sign the reports with anonymized digital identifiers that could be read by official bodies to help evaluate academic performance.

**Bernd Pulverer** *The EMBO Journal, European Molecular Biology Organization, Germany.*  
[bernd.pulverer@embo.org](mailto:bernd.pulverer@embo.org)  
*Competing interests declared* (see [go.nature.com/witfzb](http://go.nature.com/witfzb)).

## Change Chinese returnee rules

Developing countries rely on free movement of skilled scientists for the inflow of information and technology. China's rigid citizenship regulations are hindering the return of highly trained Chinese scientists from abroad, and must be changed if modernization is to be effective.

Of more than 1.62 million Chinese who left to study abroad before 2009, less than one-third have returned. China was the

second largest country of origin for science and engineering students in US higher education in 2009 (see [go.nature.com/evj2t9](http://go.nature.com/evj2t9)). Almost 90% of Chinese scientists and engineers trained overseas remained there.

At present, a Chinese researcher naturalized in another country sacrifices his or her Chinese citizenship and needs a temporary visa to return to China. Unless foreign citizenship is renounced, he or she is denied the right to open a bank account, buy a house or register a company. This bureaucracy deprives the nation of scientific and technological know-how, entrepreneurial capital, international experience and access to professional networks.

One solution would be for China to recognize a type of dual citizenship, as in India. This would allow Chinese scientists to enjoy unlimited, visa-free trips back to China and preserve such rights as access to medical care, social security, income tax and intellectual-property protection, although not the right to vote.

**Jun Li** *International Centre for Research on Environment and Development (CIRED), France.*  
[jun.li@centre-cired.fr](mailto:jun.li@centre-cired.fr)

## Worm scientist's identity revealed

The mystery scientist so hauntingly quoted on the ubiquity of roundworms in Ralph Buschbaum's 1938 textbook *Animals Without Backbones* (*Nature* **474**, 6; 2011) is biologist Nathan Cobb (1858–1932).

Cobb's pioneering work laid the foundations for the systematic discovery and study of nematodes. Members of the Nematoda are best known for supplying us with the model organism *Caenorhabditis elegans*, but it is their abundance and diversity that makes them central to biology.

Cobb would have undoubtedly been thrilled, but perhaps not surprised, by the discovery of his beloved worms more than 3 kilometres inside Earth's crust.

**Mark J. F. Brown** *Royal Holloway, University of London, Surrey, UK.*  
[mark.brown@rhul.ac.uk](mailto:mark.brown@rhul.ac.uk)

## ANIMAL BEHAVIOUR

# Born leaders

**In animals that live in groups, some individuals are leaders and others are followers. A modelling study shows that variation in leadership evolves spontaneously and need not be related to differences in knowledge or power.**

FRANZ J. WEISSING

Social animals face a dilemma. To reap the benefits of group living, they have to stay together. However, individuals differ in their preferences as to where to go and what to do next. If all individuals follow their own preferences, group coherence is undermined, resulting in an outcome that is unfavourable for everyone. Neglecting one's own preferences and following a leader is one way to resolve this coordination problem. But what attributes make an individual a 'leader'? A modelling study by Johnstone and Manica<sup>1</sup> illuminates this question.

Writing in *Proceedings of the National Academy of Sciences*, the authors consider a famous coordination problem known to game theorists as the Battle of the Sexes<sup>2</sup>. Imagine a married couple who want to spend the evening together. Husband and wife (the players) can either go to a football game or to the opera, but they cannot communicate with each other about where to meet. Neither wants to miss their partner by going to a different event from them. If that happens, both get a pay-off of zero. When they go to the same event, the wife would prefer the opera, whereas the husband would prefer the football game. When meeting at the same event, the players get the pay-offs 1 and  $1-k$  (where  $0 < k < 1$ ), depending on whether or not they realize their preferred option.

Johnstone and Manica model such an interaction (and generalize it to the case of more than two players). They assume that the same players interact repeatedly, and that each time they can either choose their preferred option or copy the previous action of the other player. Each player is characterized by a strategy,  $\lambda$ , corresponding to the player's probability of sticking to his or her preferred action. This strategy is viewed as a player's degree of leadership: players with a high value of  $\lambda$  are leaders, in that they ignore the actions of others and obey their own preferences; players with a low value of  $\lambda$  are followers, in that they tend to copy the choices of others.

Johnstone and Manica<sup>1</sup> investigate how natural selection shapes intrinsic leadership in a population in which individuals produce offspring in proportion to their pay-off in



M. JONES/MINDEN PICTURES/FLPA

**Figure 1 | Out in front.** An implication of Johnstone and Manica's model<sup>1</sup> is that leadership may simply reflect an intrinsic tendency to follow one's own preferences.

the coordination game. A population of only leaders ( $\lambda = 1$ ) is not evolutionarily stable: if both players stick to their preferred option, they will never meet and will get a pay-off of zero. Likewise, a population of only followers ( $\lambda = 0$ ) is not stable, because the players will again miss each other if both have a tendency to dither, continually switching to the previous action of the other. Instead, the population will first evolve to an intermediate value of  $\lambda$  (say,  $\lambda = 0.5$ ). But, intriguingly, this is not the final outcome. From the intermediate strategy, the population will diversify and evolve to a state where two strategies coexist — a leader strategy (say,  $\lambda = 0.9$ ) and a follower strategy (say,  $\lambda = 0.1$ ).

This outcome makes intuitive sense, because a leader–follower pair of players is most efficient in solving the coordination problem: both will eventually choose the preferred option of the leader. The leaders seem to have the better part (a pay-off of 1 is higher than a pay-off of  $1-k$ ), but this holds only when they are teamed up with a follower. On the population level,

leaders and followers have the same average pay-off. This is because leaders are more frequent than followers (because of their higher pay-off in leader–follower interactions) and therefore find themselves relatively often teamed up with another leader (giving a pay-off of zero).

These results<sup>1</sup> are interesting for several reasons. First, they provide an explanation for empirical observations in the lab and field. For example, experiments with sticklebacks<sup>3</sup> have revealed pronounced individual differences in the tendency to lead that resemble those in the model. Second, the results show that leadership and 'followership' can evolve in the absence of any other differences between individuals. In the behavioural sciences, there is much discussion about which traits make someone a leader<sup>4</sup>. According to Johnstone and Manica's model, leadership need not be associated with being better informed, being more dominant or having superior communication skills. Instead, leadership may simply reflect an intrinsic tendency to follow one's



own preferences and disregard the choices of others (Fig. 1).

The third interesting aspect is that the paper provides a link to the issue of animal 'personalities'<sup>5</sup>, the phenomenon that animals differ systematically in their behaviour in a manner that is individually stable across a variety of contexts. In nature, leadership seems to be a personality trait that is correlated with general activity level, aggressiveness and boldness<sup>3</sup>. Johnstone and Manica provide a neat explanation for the emergence of individual differences in leadership, but it is an open question how such correlations between leadership and other personality traits have evolved.

The type of model presented by Johnstone and Manica sacrifices realism for conceptual clarity and analytical tractability. It remains to be seen whether the results are robust when more-realistic assumptions are incorporated or more-complex strategies are considered. For good reason, the authors have assumed that the players do not differ in features such as knowledge and power. In more realistic settings, asymmetries between the players will undoubtedly occur; such asymmetries can help to solve a coordination problem<sup>6</sup>.

Moreover, even in symmetrical settings, differences in leadership will not necessarily evolve if more-complex strategies are available. An example can be taken directly from the authors' experimental work<sup>7</sup>: sticklebacks that have diverging preferences take turns in leadership, rather than specializing in the roles of leader and follower. Perhaps most importantly, a group of individuals engaging in prolonged interactions can be expected to learn each others' characteristics (for example, their degree of leadership). It would be worthwhile investigating how the evolutionary outcome would change if individuals could signal their leadership tendencies — as humans clearly do.

This work<sup>1</sup> may be criticized for its restricted view of leadership. One could argue that the 'leaders' in the model do not really lead, but simply refrain from following others. Leaders are defined as being stubborn, refusing to react to their fellow group members. Accordingly, the evolution of differences in leadership in this model bears some resemblance to the evolution of individual variation in responsiveness<sup>8</sup> and social sensitivity<sup>9</sup> seen in other models. In reality, there are more dimensions to leadership, and it is not obvious that stubbornness and antisocial behaviour are characteristic features of leaders. In African elephants, for example, the most responsive and socially sensitive individuals have the highest propensity to become leaders of the herd<sup>10</sup>.

Johnstone and Manica's concept of leadership seems to be most easily applicable to fish shoals and other anonymous societies. Still, even for highly structured social systems such as those of humans and elephants, their insight provides clues to how intrinsic differences in leadership could evolve as a

fundamental means to resolve the tension between individual interests and the desire to live in a group. ■

**Franz J. Weissing** is at the Centre for Ecological and Evolutionary Studies, University of Groningen, 9747 AG Groningen, the Netherlands.  
e-mail: f.j.weissing@rug.nl

1. Johnstone, R. A. & Manica, A. *Proc. Natl Acad. Sci. USA* **108**, 8373–8378 (2011).
2. Luce, R. D. & Raiffa, H. *Games and Decisions* (Wiley, 1957).

3. Harcourt, J. L., Ang, T. Z., Sweetman, G., Johnstone, R. A. & Manica, A. *Curr. Biol.* **19**, 248–252 (2009).
4. DeRue, D. S., Nahrgang, J. D., Wellman, N. & Humphrey, S. E. *Personnel Psychol.* **64**, 7–52 (2011).
5. Dingemanse, N. J. & Wolf, M. *Phil. Trans. R. Soc. Lond. B* **365**, 3947–3958 (2010).
6. Selten, R. J. *Theor. Biol.* **84**, 93–101 (1980).
7. Harcourt, J. L., Sweetman, G., Manica, A. & Johnstone, R. A. *Curr. Biol.* **20**, 156–160 (2010).
8. Wolf, M., van Doorn, G. S. & Weissing, F. J. *Proc. Natl Acad. Sci. USA* **105**, 15825–15830 (2008).
9. Wolf, M., van Doorn, G. S. & Weissing, F. J. *Proc. R. Soc. Lond. B* **278**, 440–448 (2011).
10. McComb, K., Moss, C., Durant, S. M., Baker, L. & Sayialel, S. *Science* **292**, 491–494 (2001).

## PROTEIN SYNTHESIS

# Stop the nonsense

**A subtle biochemical alteration can reprogram signals that herald the termination of protein translation into signals encoding amino acids at the level of messenger RNA — and without altering the corresponding DNA. SEE LETTER P.395**

ADRIAN R. FERRÉ-D'AMARÉ

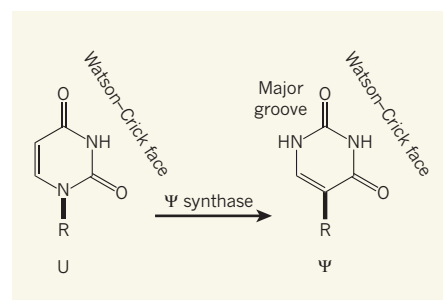
The amino-acid sequence of a protein is specified by combinations of 64 trinucleotides (or codons) in the corresponding messenger RNA. Of these, three codons, known as termination or nonsense codons, signal the end of protein translation. Sometimes, however, rather than stopping protein synthesis, the translation machinery decodes a termination codon as an amino acid in what is known as nonsense suppression. On page 395 of this issue, Karijovich and Yu<sup>1</sup> report an artificial way of inducing nonsense suppression — through post-transcriptional conversion of the uridine residue in termination codons into its isomer, pseudouridine. This finding raises fundamental questions about the biochemistry of protein synthesis and has implications for treating genetic diseases.

Translation takes place in cellular organelles called ribosomes, in which each mRNA codon is matched with the anticodon of an aminoacyl-tRNA. The latter is a transfer RNA that has been loaded by its cognate aminoacyl-tRNA-synthetase enzyme with the amino acid corresponding to its anticodon. None of the tRNAs has anticodons complementary to the termination codons; normally, proteins called release factors (RF1 and RF2 in bacteria, eRF1 in eukaryotes) recognize the nonsense codons. But if a tRNA undergoes a mutation in its anticodon such that it becomes complementary to a termination codon (and if this mutant tRNA is otherwise recognized normally by its aminoacyl-tRNA synthetase and the rest of the translation machinery), it might lead to misinterpretation of the termination codon.

Indeed, such nonsense suppression by mutated tRNAs is well documented<sup>2</sup>. The

findings of Karijovich and Yu<sup>1</sup> are surprising, however, because of their significance for the mechanism by which release factors are thought to recognize termination codons, and because of the structural similarity between pseudouridine (Ψ) and uridine (U).

The crystal structures of the bacterial ribosome with its release factors caught in the act of recognizing termination codons<sup>3,4</sup> indicate how RF1 and RF2 recognize the U of all three termination codons (UAA, UAG or UGA): chemical groups in the backbone of these release factors seem to form hydrogen bonds with groups on the face of U that normally participate in hydrogen bonding with another nucleotide — the Watson–Crick face.



**Figure 1 | Uridine and pseudouridine.** Uridine (U) — the first residue of the three termination codons — can be converted into its isomer pseudouridine (Ψ) in a reaction catalysed by pseudouridine synthase enzymes. Karijovich and Yu<sup>1</sup> show that conversion of U to Ψ can transform a termination codon into an amino-acid-coding signal. The Watson–Crick faces of U and Ψ are identical, but they differ in other details — Ψ, for instance, has an imine group (NH) that projects into the major groove of the RNA. Thick lines denote the glycosidic bond that joins the bases to the RNA backbone (R).

Although  $\Psi$  and U differ in that the former has a carbon–carbon, rather than a carbon–nitrogen, glycosidic bond and an imine (NH) group that it projects into the major groove of the RNA, the Watson–Crick faces of these two residues are identical (Fig. 1). Thus, release factors should be insensitive to conversion of the termination codons to  $\Psi$ AA,  $\Psi$ AG or  $\Psi$ GA.

What, then, is the property of  $\Psi$  — other than its ability to form Watson–Crick base pairs — that gives rise to nonsense suppression?  $\Psi$  binds water through its major-groove imine group, and this hydration makes  $\Psi$ -containing RNAs stiffer<sup>5</sup>. It could be that, when the release factors bind the  $\Psi$ -containing mRNA, the increased energy needed to dehydrate this modified mRNA results in nonsense suppression. Alternatively, nonsense suppression could be a consequence of the greater difficulty in unstacking the isomer-containing termination codon from the previous codon as the isomerized codon is brought into the ‘reading’ position on the ribosome. Regardless of the physico-chemical basis, however, the new results point to a crucial role for factors other than Watson–Crick base pairing in the recognition of termination codons.

Karijolic and Yu demonstrate nonsense suppression through pseudouridylation of termination codons both *in vitro* and in yeast. When the authors characterized the proteins synthesized following nonsense suppression, they uncovered another surprise. Rather than incorporating a random amino acid at the site occupied by the isomerized termination codon, the translation machinery specifically incorporates either serine or threonine at  $\Psi$ AA and  $\Psi$ AG, and either tyrosine or phenylalanine at  $\Psi$ GA.

This observation is noteworthy because, although the two sets of amino acids have chemical commonalities (threonine and serine both have a hydroxyl group, and tyrosine and phenylalanine share a phenyl ring), the anticodons of tRNAs for the four amino acids do not show any obvious complementarity to the termination codons. Mechanistically, this implies that pseudouridylation of termination codons leads not only to a loss of recognition by release factors, but also to a gain of recognition by specific aminoacyl-tRNAs. The fidelity of normal translation is enhanced through a proofreading process in which the accuracy of codon–anticodon pairing is communicated across the ribosome to the amino-acylated (acceptor) end of tRNA. Perhaps pseudouridylation of termination codons also affects this process.

Site-specific enzymes called pseudouridine synthases produce  $\Psi$  from U residues of cellular RNAs<sup>6</sup>. Eukaryotes and archaea have a versatile class of pseudouridine synthases called H/ACA ribonucleoproteins (RNPs)<sup>7</sup>. These complexes have their four core proteins in common, but each assembles using one of

many different RNAs (containing evolutionarily conserved sequence elements called H and ACA). The RNA component is called a guide RNA because it has a stretch of nucleotides complementary to the sequences that flank the uridine of the substrate RNA targeted for pseudouridylation. This sequence complementarity is necessary and sufficient for directing the H/ACA RNP to pseudouridylate a cellular RNA *in vivo*.

Karijolic and Yu<sup>1</sup> used a custom-designed H/ACA guide RNA to target termination codons for pseudouridylation in their yeast experiments. The authors point out that this would also be an attractive approach to treating genetic disorders that result from premature termination of translation. Indeed, more than a third of genetic disorders and many cancers are due to mutations that introduce premature termination codons<sup>8</sup>. So, rather than having to correct the mutation at the level of DNA, all that is required would be delivery of an H/ACA guide RNA to pseudouridylate the defective mRNA.

More broadly, it is possible that nature is already using this kind of ‘gene therapy’ to increase the coding capacity of genomes.

Karijolic and Yu have found several candidate mRNAs whose termination codons could be subjected to pseudouridylation by previously described H/ACA guide RNAs. Such mRNAs would produce a shorter protein in their unmodified state and a longer protein (ending at a second, unmodified termination codon) when the first termination codon is pseudouridylated. ■

**Adrian R. Ferré-D’Amaré** is at the National Heart, Lung and Blood Institute, National Institutes of Health, Bethesda, Maryland 20892-8012, USA.

e-mail: ferrea@mail.nih.gov

1. Karijolic, J. & Yu, Y.-T. *Nature* **474**, 395–398 (2011).
2. Murgola, E. J. in *tRNA Structure, Biosynthesis, and Function* (eds Söll, D. & RajBhandary, U. L.) 491–509 (ASM Press, 1995).
3. Weixlbaumer, A. *et al. Science* **322**, 953–956 (2008).
4. Laurberg, M. *et al. Nature* **454**, 852–857 (2008).
5. Charette, M. & Gray, M. W. *IUBMB Life* **49**, 341–351 (2000).
6. Hamma, T. & Ferré-D’Amaré, A. R. *Chem. Biol.* **13**, 1125–1135 (2006).
7. Hamma, T. & Ferré-D’Amaré, A. R. *J. Biol. Chem.* **285**, 805–809 (2010).
8. Frischmeyer, P. A. & Dietz, H. C. *Hum. Mol. Genet.* **8**, 1893–1900 (1999).

#### CONDENSED-MATTER PHYSICS

## Microscopy of the macroscopic

**The presence of magnetic moments in materials known as Kondo lattices can lead to an exotic transformation in their properties. The first successful endeavour into imaging such a transformation has now been made. SEE LETTER P.362**

PIERS COLEMAN

The availability of electron microscopes after the Second World War made it possible to image life at the subcellular level, prompting a revolution in cell biology. Today, a new type of microscope that images electrons at the quantum level — the scanning tunnelling microscope — is helping to drive a revolution in condensed-matter physics. But whereas biologists look on the microscope as a path to greater understanding of life at the microscopic molecular level, physicists are increasingly seeking to use it to elucidate the emergent, macroscopic properties of electrons, such as high-temperature superconductivity and magnetism. On page 362 of this issue, Ernst *et al.*<sup>1</sup> use the scanning tunnelling microscope to image the profound transformation that a metal structure known as a Kondo lattice undergoes as it dissolves a lattice of magnetic moments.

A scanning tunnelling microscope uses the

quantum-mechanical nature of electrons to image the electronic organization in metals. Although electrons lack the energy to escape from a metal, when a sharp metal tip is brought within a few ångströms of another host metal, electrons can quantum mechanically tunnel through the vacuum between the two. The tunnel current that this generates is determined by the number of available electron states in the host metal with energies up to a voltage applied between the two metals. So by measuring the current as a function of applied voltage, one obtains the detailed electron energy spectrum of the host metal at the tip location (Fig. 1).

Modern scanning tunnelling microscopes allow one to scan the tip across the metal, gradually building up a spatial profile of the electron energy spectrum with subångström resolution. The development of the control and isolation technology necessary to hold a tunnelling tip with subångström fidelity, even as the temperature is varied, has taken several decades of development, but today the



approach is poised to open up a whole generation of electronic media to levels of scrutiny that, until recently, could only be imagined.

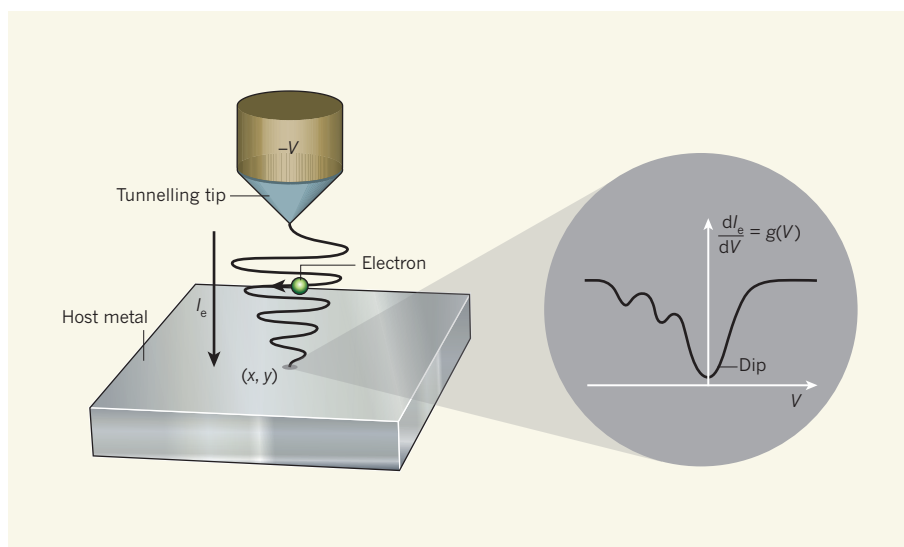
One of the long-standing interests of condensed-matter physicists concerns the interaction of magnetism with electrons. Magnetism is driven by unpaired electrons that have become localized in an atomic orbital to form tiny magnets, or magnetic moments. Materials with partially filled f- and d-orbitals are particularly prone to forming magnetic moments. The high-powered rare-earth neodymium magnets used inside the motor of a modern hybrid car, for example, rely on the magnetism associated with f-electrons.

Sometimes, such magnetic moments can actually dissolve inside a metal, profoundly modifying the metal's properties, driving up the effective mass of the mobile electrons by several orders of magnitude to form 'heavy electrons'. This situation can be likened to the behaviour of ions in water, in which the electrically polar water molecules 'screen' the ions and dissolve them into the solution. In the course of this process, the ions become mobile, and change the conductivity and chemistry of the solution. Similarly, in a metal the magnetically polar electron fluid tends to screen local magnetic moments immersed in it. During this process, the spin originally on the magnetic moment delocalizes as a charged heavy electron.

But whereas the solvation of ions is thermally driven — improving at high temperatures — the screening of magnetic moments in a metal is a quantum process. For the magnetic moment to be screened, it needs to become quantum mechanically entangled with the magnetic moment of the surrounding electrons. Entanglement requires that the quantum-mechanical phase difference of the electrons involved remains well defined, and this occurs only at low temperatures, in which the 'dephasing' effects of temperature are negligible.

The screening of magnetic moments inside a metal is called the Kondo effect after its discoverer, Jun Kondo. Several years after Kondo's pioneering work<sup>2</sup>, Sebastian Doniach<sup>3</sup> showed that the same physics can occur in materials containing a dense array of magnetic moments immersed in a metal, called Kondo lattice materials. In these materials, the disappearance of the magnetic moments into the electron fluid profoundly changes the bulk properties, lowering the resistivity of the metal, and sometimes producing superconductivity.

In their study, Ernst *et al.*<sup>1</sup> describe the first successful attempt to image the magnetic screening process in a Kondo lattice material using scanning tunnelling electron microscopy (STM). The observation of the Kondo effect at single magnetic ions was one of the early successes of STM, but it is only in the past year that the first successful STM measurements were carried out<sup>4,5</sup> on an f-electron metal,



**Figure 1 | Electron tunnelling from tip to host.** Application of a voltage ( $-V$ ) to the tunnelling tip causes a current of electrons ( $I_e$ ) to tunnel through the vacuum into the host metal. The derivative of the current ( $dI_e/dV$ ) determines the energy spectrum  $g(V)$  of the metal as a function of voltage;  $(x, y)$  denotes the coordinates of the point at which the current enters the metal. Ernst *et al.*<sup>1</sup> observe a crater-like dip in the energy spectrum around each ytterbium site in the ytterbium rhodium material  $\text{YbRh}_2\text{Si}_2$ .

the uranium ruthenium compound  $\text{URu}_2\text{Si}_2$ . However,  $\text{URu}_2\text{Si}_2$  is complicated by the multi-electron physics of the uranium ion. Building on this recent success, Ernst *et al.*<sup>1</sup> chose an ytterbium rhodium material,  $\text{YbRh}_2\text{Si}_2$ , which has the same structure as  $\text{URu}_2\text{Si}_2$  but in which the local magnetic moments form an unambiguous Kondo lattice.

In their single-atom-resolution STM images, Ernst *et al.*<sup>1</sup> observe a crater-like dip in the tunnelling current around each ytterbium site (Fig. 1). This feature is a classic interference signature of the Kondo effect. The interference causes the STM signal to die out at the location of the magnetic moments. And this is precisely what is seen: as the temperature is lowered and the magnetic moments become entangled with the system's conduction electrons, a dip develops in the material's electronic density of states. Moreover, at an applied voltage of 6 millivolts, the authors see a small feature that develops below 30 kelvin, which they identify as the resonant state associated with the Kondo effect. Unlike previous work on single magnetic moments, here the authors see these signatures in every unit cell of the lattice on the surface of the material, establishing the dense nature of the magnetic screening.

A second important feature of the new results<sup>1</sup> is the observation of a magnetic effect known as crystal-field excitations in an STM measurement. In a crystalline environment, the energy levels of a magnetic moment are split into several crystal-field excitations. Crystal-field excitations are neutral and don't directly couple to electron currents, yet Ernst and colleagues observe them directly in the electronic density of states — a definite indication that, through screening, the magnetic ions have become coupled to the mobile electrons.

Moreover, the crystal-field excitations are in precisely the right position as measured by neutron-scattering experiments. The authors are able to confirm this interpretation through detailed calculations on the Kondo effect in a crystalline lattice environment.

Perhaps the most exciting aspect of these measurements<sup>1</sup> is the prospects they hold for the future. The current measurements have all been made at above 4.5 kelvin, but at lower temperatures the quantum fluctuations associated with the Kondo lattice become so intense that the masses of the heavy electrons diverge, leading to a 'quantum critical point' at which the heavy electrons are thought to break up into magnetic spins<sup>6</sup>. At a critical point, the fluctuations expand to macroscopic dimensions — something that can be seen in the opalescence of superheated water at the critical point. The analogous quantum criticality, involving quantum fluctuations of macroscopic dimensions, has never been observed directly. The excitement of Ernst and colleagues' results lies in the distinct possibility that such macroscopic criticality in the quantum waves of the Kondo lattice might soon be directly imaged by STM. ■

**Piers Coleman** is in the Center for Materials Theory, Department of Physics and Astronomy, Rutgers University, New Jersey 08854-8019, USA.  
e-mail: coleman@physics.rutgers.edu

- Ernst, S. *et al.* *Nature* **474**, 362–366 (2011).
- Kondo, J. *Prog. Theor. Phys.* **32**, 37–49 (1964).
- Doniach, S. *Physica B* **91**, 231–234 (1977).
- Schmidt, A. R. *et al.* *Nature* **465**, 570–576 (2010).
- Aynajian, P. *et al.* *Proc. Natl Acad. Sci. USA* **107**, 10383–10388 (2010).
- Custers, J. *et al.* *Nature* **424**, 524–527 (2003).

## STRUCTURAL BIOLOGY

# Porthole to catalysis

The crystal structure of a sugar-transferring enzyme offers insight into the mechanism of a ubiquitous protein-modification reaction, and solves the mystery of how the enzyme recognizes certain sequences in proteins. [SEE ARTICLE P.350](#)

REID GILMORE

One of the most common protein-modification reactions in the cells of eukaryotes (organisms that include plants, animals and fungi) is *N*-linked glycosylation, in which sugars are attached to the side chain of the amino acid asparagine. This reaction has diverse roles in protein folding and stability, intracellular trafficking and cell-cell interactions, and is catalysed by the enzyme oligosaccharyltransferase (OST). More specifically, OST mediates the transfer of an oligosaccharide from a donor substrate onto acceptor asparagine residues in newly synthesized proteins. In eukaryotes, such

glycosylation occurs only at asparagine residues located within asparagine-X-threonine/serine amino-acid sequences (N-X-T/S sequences), where X can be any amino acid except proline. It has been unclear how OST recognizes the acceptor sites and the associated amino-acid sequences, and how it activates the normally unreactive nitrogen in asparagine's side chain to take part in glycosylation. But on page 350 of this issue, Lizak *et al.*<sup>1</sup> now provide remarkable insight into these issues with their report of the X-ray crystal structure of PglB, an OST from the bacterium *Campylobacter lari*.

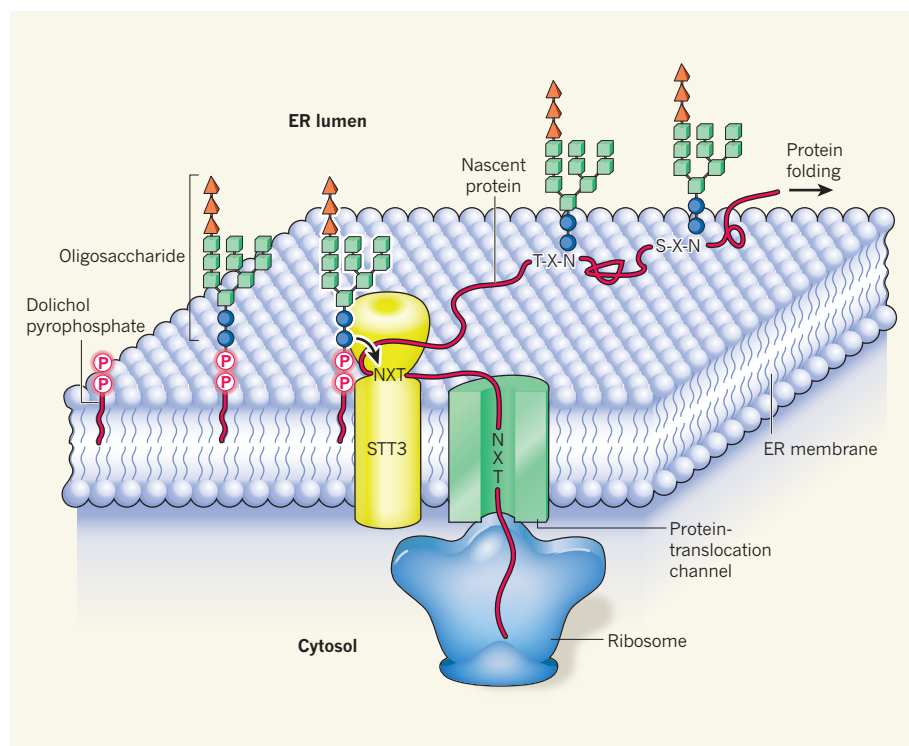
Although often described as post-translational modifications, most *N*-linked oligosaccharides in eukaryotic cells are added

co-translationally to a nascent polypeptide as it is threaded through the protein-translocation channel in the endoplasmic reticulum (ER) membrane<sup>2</sup> (Fig. 1). The preassembled oligosaccharide is anchored by a lipid to the luminal surface of the ER. Consequently, the OST has to scan nascent polypeptides for acceptor sites that are moving past the enzyme at the rate of protein synthesis (about 6–8 residues per second for eukaryotes), while simultaneously positioning the lipid-linked oligosaccharide (LLO) in the enzyme's active site.

Unlike the oligomeric OST complexes that are present in eukaryotes<sup>3</sup>, PglB is monomeric, making it a more suitable candidate for protein crystallography. PglB consists of an amino-terminal domain that binds to the inner bacterial membrane, and a carboxy-terminal domain that resides in the periplasm (the space between the inner and outer bacterial membranes). Proteins are glycosylated by PglB as they thread through the plasma membrane. The enzyme's periplasmic domain includes a short, evolutionarily conserved amino-acid sequence (tryptophan-tryptophan-aspartic acid, abbreviated as WWD) that is essential in all OST catalytic subunits<sup>4,5</sup>. The structure of this domain in PglB from the archaeon *Pyrococcus furiosus* has been solved previously<sup>6</sup>, but the structure of the membrane-bound domain was unknown, not least because crystallizing integral membrane proteins is always technically challenging. Lizak *et al.*<sup>1</sup> surmounted this challenge by crystallizing the complete *Campylobacter* PglB protein in the presence of its peptide substrate.

The authors' structure reveals that two large loops (EL1 and EL5) of the N-terminal domain extend into the periplasm to form a platform that supports the periplasmic domain, and contain residues required for peptide binding and catalysis. Remarkably, the peptide-binding cleft is formed by the interface between the membrane-binding and periplasmic domains, and is located on the opposite face of PglB from the cleft that harbours the catalytic site and the LLO-binding site. This architecture explains how the large LLO and peptide substrates can independently enter the enzyme's active site. The structure is consistent with biochemical studies<sup>7</sup> that indicated that acceptor sequences must be located within flexible or unfolded segments of polypeptides. It also shows that the critical WWD motif binds to the threonine side chain of the peptide substrate, thereby explaining why serine or threonine must be located two amino acids after asparagine in the N-X-T/S acceptor sequence.

Another intriguing feature of the structure<sup>1</sup> is that the asparagine side chain in the peptide projects through a narrow 'porthole' of PglB that connects the peptide-binding cleft to the catalytic site. This site is formed by a cluster of charged residues that bind a magnesium ion ( $Mg^{2+}$ ). Although OST activity has long been known to be dependent on a divalent cation



**Figure 1 | Oligosaccharide transfer reactions.** In eukaryotic cells, proteins made by the ribosome are passed across the endoplasmic reticulum (ER) membrane through a protein-translocation channel. Oligosaccharyltransferase (OST) proteins adjacent to the channel catalyse the attachment of oligosaccharides to the proteins in *N*-glycosylation reactions before the proteins acquire secondary or tertiary structure. For simplicity, only the active-site subunit of a eukaryotic OST (known as STT3) is depicted. The oligosaccharides are assembled on lipid carriers (dolichol pyrophosphate molecules; P is a phosphate group) in the membrane, and are attached to asparagine (N) side chains within N-X-T/S acceptor sequences, where T is threonine, S is serine and X is any amino acid other than proline. Lizak and colleagues' crystal structure<sup>1</sup> of PglB (a bacterial protein analogous to STT3 in eukaryotes) in complex with an acceptor peptide suggests how OST active-site subunits scan nascent polypeptides for N-X-T/S sequences while positioning the lipid-linked oligosaccharide in the catalytic site.



(either  $\text{Mg}^{2+}$  or the manganese ion,  $\text{Mn}^{2+}$ ), the catalytic-site residues had not been identified. Lizak *et al.* report that three of the active-site residues have acidic side chains (they are aspartic acid or glutamic acid residues). When the authors replaced any of these residues with alanine, which has a non-acidic, methyl side chain, the glycosylation activity of the resulting enzymes was reduced by 50–90% compared with the wild-type enzyme.

The side chain of asparagine contains an amide group ( $\text{CONH}_2$ ) in which the nitrogen atom is a 'weak nucleophile' — which means that it shouldn't react readily with the oligosaccharide of the LLO. So how does the enzyme activate the amide so that it can react? Previous models for OST catalysis envisaged that activation occurred through the formation of hydrogen bonds between the asparagine side chain and the threonine (or serine) residue in the peptide substrate<sup>8,9</sup>. But Lizak and colleagues' finding<sup>1</sup> that the asparagine residue projects through the porthole in the active site rules out this possibility.

Instead, the authors propose that two active-site residues form hydrogen bonds to the amide hydrogens of asparagine's side chain, and in so doing increase the ability of the amide to react with the oligosaccharide. When the researchers replaced these catalytic residues with others to alter the hydrogen bonding to the asparagine side chain, the resulting mutants lost their catalytic activity, thus providing strong support for the proposed scheme. This activation mechanism could also explain why a glutamine-containing pseudo-acceptor sequence (Q–X–T/S, where Q is glutamine) can bind to the OST active site, yet not be glycosylated — the glutamine side chain cannot form properly positioned hydrogen bonds to the active site<sup>10</sup>.

Following glycosylation, the asparagine side chain on the polypeptide substrate is covalently linked to a bulky oligosaccharide that cannot pass through the narrow active-site porthole. So how can the reaction product leave the active site? Lizak and colleagues' structure<sup>1</sup> reveals that the porthole is formed by the packing of the flexible, partially disordered EL5 loop against the periplasmic domain. They therefore propose that product formation induces a conformational change that promotes disengagement of EL5 from the periplasmic domain, followed by release of the glycosylated product and the lipid anchor.

Lizak and colleagues' crystal structure of PglB in complex with a peptide substrate provides invaluable information about PglB–peptide binding and the enzyme's catalytic mechanism, but higher-resolution structures are needed to provide detailed insight into the mechanism. What's more, the location of the LLO binding site in PglB cannot be precisely known until the structure of a PglB–LLO complex is solved. Such a structure should also reveal why *N*-acetylglucosamine (a sugar that

contains an  $\text{NHCOCH}_3$  group in place of one OH group) is present at the reducing end of all eukaryotic LLO donors — that is, at the end that connects the oligosaccharide to the lipid<sup>11</sup>. Finally, the structure of an enzyme–product complex would provide insight into how the product dissociates from the enzyme. ■

**Reid Gilmore** is in the Department of Biochemistry and Molecular Pharmacology, University of Massachusetts Medical School, Worcester, Massachusetts 01605, USA.  
e-mail: reid.gilmore@umassmed.edu

1. Lizak, C., Gerber, S., Numao, S., Aebi, M. & Locher, K. P. *Nature* **474**, 350–354 (2011).

2. Ruiz-Canada, C., Kelleher, D. J. & Gilmore, R. *Cell* **136**, 272–283 (2009).
3. Kelleher, D. J. & Gilmore, R. *Glycobiology* **16**, 47R–62R (2006).
4. Wacker, M. *et al. Science* **298**, 1790–1793 (2002).
5. Yan, Q. & Lennarz, W. J. *J. Biol. Chem.* **277**, 47692–47700 (2002).
6. Igura, M. *et al. EMBO J.* **27**, 234–243 (2008).
7. Kowarik, M. *et al. Science* **314**, 1148–1150 (2006).
8. Imperiali, B., Shannon, K. L., Unno, M. & Rickert, K. W. *J. Am. Chem. Soc.* **114**, 7944–7945 (1992).
9. Bause, E. & Legler, G. *Biochem. J.* **195**, 639–644 (1981).
10. Nilsson, I. *et al. J. Cell Biol.* **161**, 715–725 (2003).
11. Wacker, M. *et al. Proc. Natl Acad. Sci. USA* **103**, 7088–7093 (2006).

## ASTROPHYSICS

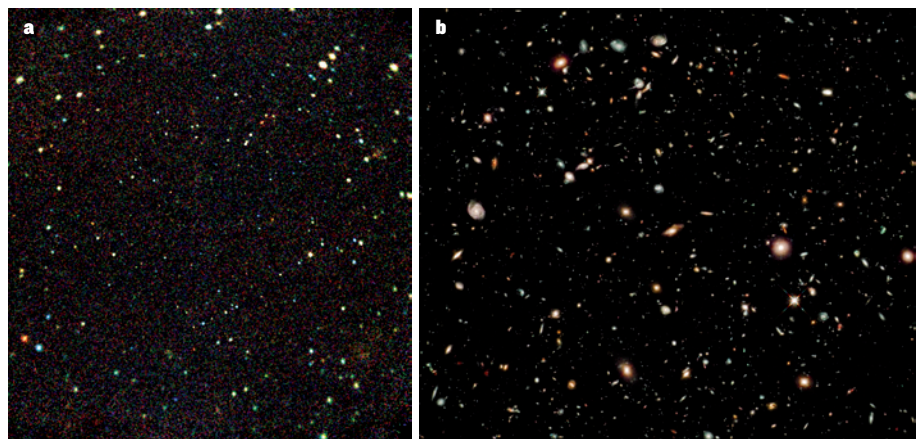
## Early black holes uncovered

**Collective X-ray emission from distant galaxies reveals a hidden population of supermassive black holes. This finding suggests that galaxies and their central black holes have been coevolving since early cosmic times. SEE LETTER P.356**

ALEXEY VIKHLININ

The formation of the first stars and black holes a few hundred million years after the Big Bang marks the end of the cosmic 'dark ages'<sup>1</sup>. These objects heat and ionize the surrounding gas, and by a billion years after the Big Bang (or redshift ( $z$ ) of about 6) they have ionized nearly all<sup>2</sup> of the hydrogen in the Universe. This grand picture is mostly a product of theoretical modelling — data on the first generations of objects are scarce. Only recently have candidate galaxies at  $z$  higher than 6 been

identified in deep Hubble Space Telescope images<sup>3</sup>. What about the supermassive black holes that presumably lie at their centres? A small number of quasars (extremely luminous galactic nuclei powered by supermassive black holes) at  $z$  of about 6 have been discovered in the Sloan Digital Sky Survey<sup>4</sup>. But these objects are so luminous and massive that they should represent only the very tip of the iceberg of the overall black-hole population. On page 356 of this issue, Treister *et al.*<sup>5</sup> present a detection of signal from much more 'typical' black holes, at  $z$  about 6, found quite literally



**Figure 1 | Teaming up Chandra with Hubble.** By comparing and cross-correlating a Chandra X-ray Observatory image of a patch of the sky known as Deep Field South (a) with observations of the same area taken by the Hubble Space Telescope (b), Treister and colleagues<sup>5</sup> discovered a population of supermassive black holes at early cosmic times.

A. NASA/JHU/UM/R. GIACCONI ET AL.; B. NASA, ESA, G. ILLINGWORTH & R. BOUWENS (UNIV. CALIFORNIA, SANTA CRUZ), THE HUDF09 TEAM

by sorting through the X-ray photons in the deepest images taken by the Chandra X-ray Observatory.

X-rays are the ubiquitous observational signature of gas falling into black holes. X-ray emission more easily penetrates dense gas and dust clouds than does optical or ultraviolet (UV) light, and so it is considered to be one of the cleanest methods of finding active black holes. Chandra's image of the 'Deep Field South' patch (Fig. 1) indeed contains hundreds of active black holes, mostly in quasars at  $z = 1-3$  (ref. 6). However, no black holes at  $z$  greater than 6 have hitherto been found in X-ray observations. The Chandra images are too small to contain rare 'monsters' such as the Sloan quasars, and are not sensitive enough to detect more typical supermassive black holes at these high redshifts. However, we know from Hubble observations of the same field that, in the area between detected Chandra sources, there are hundreds of galaxies at  $z$  higher than 6.

In their study, Treister *et al.*<sup>5</sup> cross-correlated the Chandra X-ray photons with the locations of the Hubble galaxies and found a positive signal. The observed X-ray signal is very low — less than five X-ray photons per galaxy — and is detectable only because it was averaged over almost 200 high- $z$  galaxies, yielding an effective exposure time of 23 years. But the statistical significance of this detection is high (nearly 7 sigma, to use the technical jargon), so we have a confident detection of X-ray emission from a population of typical supermassive black holes hosted by high- $z$  galaxies.

A far-reaching finding of this analysis<sup>5</sup> is the X-ray 'colours' of these high- $z$  galaxies. The observed X-ray emission spectrum peaks at energies greater than 3 kiloelectronvolts (keV). Because the Universe's expansion stretches an object's light waves by a factor of  $1 + z$ , this corresponds to a spectral peak at energies greater than 20 keV in the galaxies' reference frame. This observation strongly indicates that the central black holes in nearly all high- $z$  galaxies within the Chandra field are blanketed in large amounts of cold gas that absorbs softer (lower-energy) X-rays. This same gas would completely absorb the optical and UV emission, making these black holes detectable only in the X-ray band.

The authors estimate the total mass of the newly discovered black-hole population using the notion<sup>7</sup> that about 10% (within a factor of a few) of rest energy of infalling matter is radiated away, and by further assuming that the black holes have been accreting at the same rate for a substantial fraction of the age of the Universe at their redshift. A limiting factor in such an estimate is that only a small fraction of the black-hole luminosity is in the hard (high-energy) X-ray regime directly observed by Chandra. Another caveat is that more than 95% of the black holes' luminosity, emitted mostly in the UV band, never reaches

outside the galaxies because of absorption.

The reader might therefore conclude that the authors' estimate of the total mass of the high- $z$  black-hole population, which is based solely on Chandra's X-ray emission, is highly uncertain, even though such a high uncertainty is usually unavoidable for ground-breaking astronomical observations. With these reservations in mind, it is still intriguing that Treister *et al.* conclude that the estimated black-hole masses are consistent with a hypothesis in which the relationship between galaxy mass and black-hole mass that is observed in the local Universe<sup>8,9</sup> is already established a billion years after the Big Bang.

Treister and colleagues' results<sup>5</sup> have implications for many studies of the early Universe. Unfortunately, however, answers to some key questions — such as how the progenitors of these early supermassive black holes were generated, or the exact mechanism that underlies the coevolution of the black holes and their host galaxies — will probably have to wait for the next generation of telescopes. These telescopes should be capable of detecting the

objects individually in the X-ray band and of observing their absorbed and re-radiated emission in the sub-millimetre and far-infrared regimes. But Treister *et al.* have taken the first, most difficult, step: the detection of a typical population of supermassive black holes near the end of the cosmic dark ages. ■

**Alexey Vikhlinin** is in the High Energy Astrophysics Division, Harvard-Smithsonian Center for Astrophysics, Cambridge, Massachusetts 02138, USA.  
e-mail: avikhlinin@cfa.harvard.edu

1. Bromm, V., Yoshida, N., Hernquist, L. & McKee, C. F. *Nature* **459**, 49–54 (2009).
2. Gunn, J. E. & Peterson, B. A. *Astrophys. J.* **142**, 1633–1641 (1965).
3. [www.spacetelescope.org/news/heic1001](http://www.spacetelescope.org/news/heic1001)
4. Fan, X. *et al. Astron. J.* **132**, 117–136 (2006).
5. Treister, E., Schawinski, K., Volonteri, M., Natarajan, P. & Gawiser, E. *Nature* **474**, 356–358 (2011).
6. Szokoly, G. P. *et al. Astrophys. J. Suppl. Ser.* **155**, 271–349 (2004).
7. Soltan, A. *Mon. Not. R. Astron. Soc.* **200**, 115–122 (1982).
8. Gebhardt, K. *et al. Astrophys. J.* **539**, L13–L16 (2000).
9. Ferrarese, L. *Astrophys. J.* **578**, 90–97 (2002).

#### GENE EXPRESSION

## The autism disconnect

**Separating primary from secondary changes in the autistic brain has long been a research goal. With knowledge of wide-ranging molecular deficits, identification of the best therapeutic targets becomes a priority. SEE LETTER P.380**

ŽELJKA KORADE & KÁROLY MIRNICS

**A**utism spectrum disorder is a complex neurodevelopmental condition. It is characterized by altered social interactions, communication difficulties and repetitive patterns of behaviour. There is no known single cause of autism, but it is believed that genetic predisposition together with environmental influences lead to molecular changes in brain cells, altering normal brain development<sup>1</sup>. On page 380 of this issue, Voineagu *et al.*<sup>2</sup> present the first appropriately powered and comprehensive gene-expression analysis of autistic brains using cutting-edge technologies and excellent data-mining approaches.

The authors measured messenger RNA levels for more than 30,000 genes in three regions of post-mortem brains from 19 patients with autism and 17 controls. They identified 444 genes that were differentially expressed between the cerebral cortices of the autistic and control brains. They then replicated most of the data with a second, independent cohort of post-mortem brains.

Brain regions differ in cellular composition, connectivity and molecular signatures, which, together, lead to functional specialization. In

humans, for example, the prefrontal cortex is primarily responsible for higher cognitive processes such as working memory, whereas the temporal cortex is crucial for auditory perception and semantics. Voineagu *et al.* report that the differential patterns of gene expression that normally distinguish the frontal and temporal cortices are significantly attenuated in the autistic brain. This disappearance of differential gene expression — which may also occur in other regions of the autistic brain — is very intriguing. Loss of cortical patterning may impair connectivity between the brain regions and, ultimately, weaken the specialized functions of the cortical areas.

To investigate the functional relationship between gene-expression changes, the authors<sup>2</sup> used two distinct data-mining methods. One method groups genes together on the basis of their known cellular, molecular or functional characteristics<sup>3</sup>; the other builds functional gene modules according to the observed co-expression relationships<sup>4</sup>. The authors identified two distinct gene-expression modules associated with autism that might act in concert to disrupt typical brain development. One of the modules, which mediates synaptic communication between neurons, showed



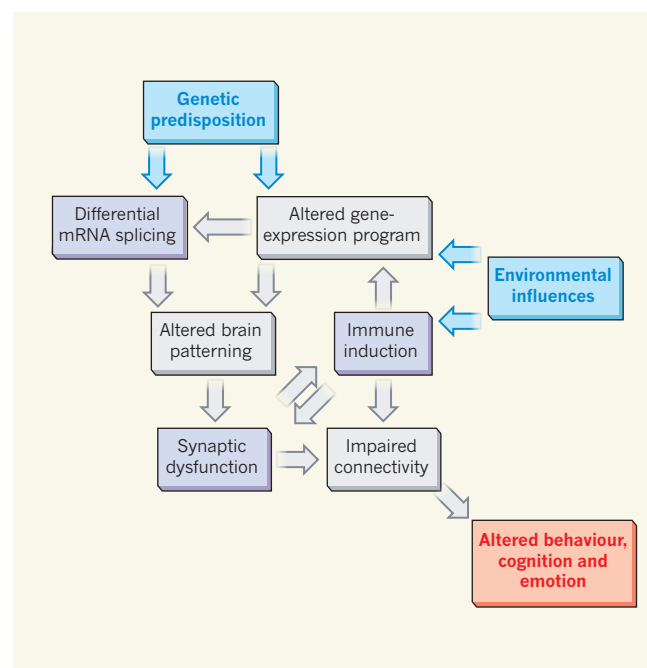
decreased expression in the autistic cortices. The other was related to immune-system activation in the brain and showed increased expression. These observations are consistent with previous findings<sup>5,6</sup> that the brains of patients with autism show altered expression of genes relating to normal synapse development and increased expression of genes mediating inflammation.

The central gene in the synaptic module was *A2BP1*, which functions in a pathway that has been implicated<sup>7</sup> in autism. The *A2BP1* protein controls what portions of a gene are included in the mature, functional mRNA during the process of alternative splicing. Through genome-wide assessment of autistic brains with decreased *A2BP1* expression, Voineagu *et al.* found *A2BP1*-dependent deficits in the RNA-splicing assembly of many genes. So it is possible for reduced expression of a single gene to affect a huge number of other genes that are responsible for normal brain development. Because diminished patterning of the cortex might arise from altered brain development, one could hypothesize that a general developmental dysregulation of transcript splicing leads to a more uniform development of various cortical areas, preventing proper functional specialization of these brain regions. This is an interesting and little-studied concept.

Are the gene-expression differences reported here primary (genetic) or environmentally induced? In search of an answer, the researchers<sup>2</sup> analysed DNA data from a published genome-wide association study of autism<sup>8</sup> to test whether expression changes in the genes of synaptic and immune modules were due to specific sequences in the patients with autism. Compared with controls, genes that were involved in synaptic function (and showed altered RNA expression in post-mortem samples) were also strongly associated with a genetic predisposition to autism, suggesting that these differences are probably strong contributors to the development of this disorder.

But perhaps the authors' most intriguing finding concerns the immune module. Immune dysfunction has been suggested<sup>6</sup> to occur in autism, although not by an unbiased genome-wide assessment. Voineagu *et al.* now show that — unlike expression changes in synapse-related genes, which seem to be due to genetic predisposition to autism — the immune response of the autistic brain is probably a non-genetic, adaptive or environmental process.

These data should, however, be interpreted cautiously. And they do not diminish the



**Figure 1 | The complexity of autism.** In autism, genetic predisposition and environmental influences disrupt typical gene-expression patterns; this, in turn, alters brain development. Voineagu and colleagues<sup>2</sup> suggest that whereas alterations in messenger RNA splicing and synaptic disturbances are primarily controlled by genetics, immune responses in the autistic brain are either adaptive or environmental. Together, these deficits cause impaired connectivity — and, ultimately, altered behaviour, cognition and emotion.

importance of immune-system activation in autism as a therapeutic target. Classification of primary and secondary changes in complex brain disorders is somewhat artificial, and one cannot be certain which set of changes mainly contributes to the symptoms. But this is not necessarily a problem. After all, some of the most common therapies — those that treat fever in influenza, cough in upper respiratory infection or pain in arthritis — target disease symptoms (and not causes), providing relief for the patients. Besides, owing to high genetic predisposition and heterogeneity among patients, causal treatments for autism might not be possible in the foreseeable future, whereas therapies targeting the common, symptom-related molecular pathways might be within reach.

But how can we establish which gene-expression alterations are most critical and how they relate to the symptoms of autism? Is a defect in alternative splicing more important than disturbances in synaptogenesis or immune-system induction? Can we even assess them independently, or are the various impairments intertwined? Unlike our notable ability to uncover autism-related molecular/genetic changes, our ability to make clinical sense of the observed and validated changes is limited. Linking molecular abnormalities to disease symptoms and manifestations is particularly challenging in the absence of true animal models — as is the case for autism.

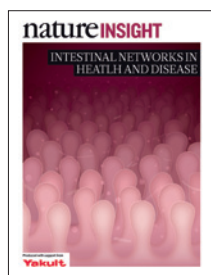
One of the strengths of Voineagu and colleagues' study<sup>2</sup> is that it provides a framework for a testable model (Fig. 1). The main priority is to determine the extent and origin of the differential splicing in the autistic brain, and its effect on the development of various brain regions. We must also establish whether splicing dysregulation is limited to *A2BP1*-targeted genes alone or is more widespread. The causality of the various changes is another fascinating issue: do the genetics-driven, converging synaptic alterations activate a detrimental immune response, or does the immune response have a more pronounced and potentiated effect when the synaptic genes show genetic vulnerability? As the authors suggest, how the changes they report relate to other neurodevelopmental disorders such as schizophrenia and attention deficit hyperactivity disorder should also be tested, given the emerging evidence<sup>9</sup> of genetic overlap.

Finally, the molecular changes ought to be correlated with the core features of the disease. The available data suggest that autism-

related genetic and molecular changes are not present in all patients. However, there has not been a combined molecular-behavioural classification of autism spectrum disorder into subgroups — such as 'splicing' autism, 'synaptic' autism and 'inflammatory' autism. Should such a molecular classification become achievable (and linkable to specific symptoms of the disorder), it would revolutionize autism research, and open the door to developing more targeted and individualized therapies. ■

**Željka Korade and Károly Mirnics** are in the Department of Psychiatry and in the Vanderbilt Kennedy Center for Research on Human Development, Vanderbilt University, Nashville, Tennessee 37232, USA.  
e-mails: [zeljka.korade@vanderbilt.edu](mailto:zeljka.korade@vanderbilt.edu); [károly.mirnics@vanderbilt.edu](mailto:károly.mirnics@vanderbilt.edu)

1. Geschwind, D. H. & Levitt, P. *Curr. Opin. Neurobiol.* **17**, 103–111 (2007).
2. Voineagu, I. *et al.* *Nature* **474**, 380–384 (2011).
3. Dennis, G. *et al.* *Genome Biol.* **4**, P3 (2003).
4. Langfelder, P. & Horvath, S. *BMC Bioinformatics* **9**, 559 (2008).
5. Judson, M. C., Eagleson, K. L. & Levitt, P. *J. Neurodev. Disord.* doi:10.1007/s11689-011-9081-8 (2011).
6. Garbett, K. *et al.* *Neurobiol. Dis.* **30**, 303–311 (2008).
7. Martin, C. L. *et al.* *Am. J. Med. Genet. B* **144B**, 869–876 (2007).
8. Wang, K. *et al.* *Nature* **459**, 528–533 (2009).
9. Williams, N. M. *et al.* *Lancet* **376**, 1401–1408 (2010).



Cover illustration by  
Nik Spencer

**Editor, *Nature***  
Philip Campbell

**Publishing**  
Nick Campbell

**Insights Editor**  
Ursula Weiss

**Production Editor**  
Nicola Bailey

**Art Director**  
Kelly Buckheit Krause

**Art Editor**  
Nik Spencer

**Sponsorship**  
Gerard Preston

**Production**  
Emilia Orviss

**Marketing**  
Elena Woodstock,  
Hannah Phipps

**Editorial Assistant**  
Hazel Mayhew

Our bowels have two major roles: the digestion and absorption of nutrients, and the maintenance of a barrier against the external environment. They fulfil these functions in the context of, and with help from, tens of trillions of resident microbes, known as the gut microbiota.

This Insight has as its topic the various relationships that contribute to keeping this complex system in balance, and that help to protect us from a wide spectrum of diseases, including chronic inflammatory bowel disease, colorectal cancer and metabolic disease. It explores the interactions between the environment and host genetics, between the type and amount of food we eat and the composition of the microbial community, and between the microbiota, the intestinal epithelium and the immune system. It also highlights the regulatory mechanisms that control the rapid and continuous renewal of epithelial cells in the intestinal lining from resident stem cells. It discusses the cellular and molecular pathways that help to maintain intestinal homeostasis, and explores the mechanisms that cause pathology and disease when these pathways fail.

We hope that these articles will contribute to a better understanding of the nature of these complex networks, and point to future strategies that will be successful in paving the way for more effective preventative and therapeutic measures, ultimately benefitting human health.

We are pleased to acknowledge the financial support of Yakult Honsha in producing this Insight. As always *Nature* carries sole responsibility for all editorial content and peer review.

**Ulla Weiss**  
*Insights Editor*

### CONTENTS

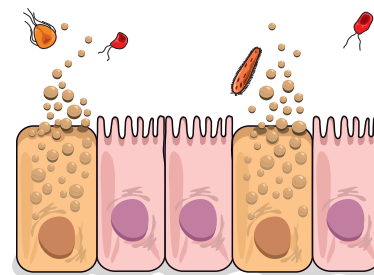
#### REVIEWS

##### 298 Intestinal homeostasis and its breakdown in inflammatory bowel disease

Kevin J Maloy & Fiona Powrie

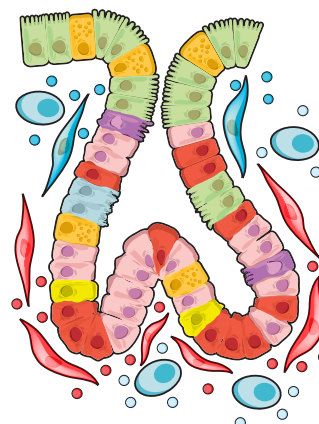
##### 307 Genetics and pathogenesis of inflammatory bowel disease

Bernard Khor, Agnès Gardet & Ramnik J Xavier



##### 318 Microenvironmental regulation of stem cells in intestinal homeostasis and cancer

Jan Paul Medema & Louis Vermeulen



#### PERSPECTIVE

##### 327 Human nutrition, the gut microbiome and the immune system

Andrew L Kau, Philip P Ahern, Nicholas W Griffin, Andrew L Goodman & Jeffrey I Gordon

The Macmillan Building  
4 Crinan Street  
London N1 9XW, UK  
Tel: +44 (0) 20 7833 4000  
e: nature@nature.com



nature publishing group



# Intestinal homeostasis and its breakdown in inflammatory bowel disease

Kevin J. Maloy<sup>1</sup> & Fiona Powrie<sup>1,2</sup>

**Intestinal homeostasis depends on complex interactions between the microbiota, the intestinal epithelium and the host immune system. Diverse regulatory mechanisms cooperate to maintain intestinal homeostasis, and a breakdown in these pathways may precipitate the chronic inflammatory pathology found in inflammatory bowel disease. It is now evident that immune effector modules that drive intestinal inflammation are conserved across innate and adaptive leukocytes and can be controlled by host regulatory cells. Recent evidence suggests that several factors may tip the balance between homeostasis and intestinal inflammation, presenting future challenges for the development of new therapies for inflammatory bowel disease.**

Inflammatory bowel disease (IBD) refers to chronic inflammatory disorders that affect the gastrointestinal tract<sup>1</sup>. There are two main clinical forms of IBD — Crohn's disease, which can affect any part of the gastrointestinal tract, and ulcerative colitis, in which pathology is restricted to the colonic mucosa<sup>1</sup>. The precise aetiology of IBD remains unclear, but several factors that make a major contribution to disease pathogenesis have been identified<sup>1</sup>. These fall into three distinct categories: genetic factors, the host immune system, and environmental factors such as the gut microbiota, which is dominated by intestinal bacteria<sup>2</sup>.

On a cellular level, the dynamic crosstalk between intestinal epithelial cells (IECs), intestinal microbes and local immune cells represents one of the fundamental features of intestinal homeostasis<sup>3,4</sup>. These interactions are not only important for the pathogenesis of IBD, but also essential for maintaining normal intestinal homeostasis and for mounting protective immunity to pathogens. In this Review, we summarize recent findings from disease models and clinical samples that illustrate key interactions and pathways that regulate intestinal homeostasis. We discuss how defects in epithelial barrier function, innate immune recognition or immune regulatory circuits may precipitate the aberrant expression of pathological inflammatory responses in IBD. Finally, we offer some perspectives on future challenges for developing therapies for IBD.

## Regulation by the epithelial barrier

The intestinal epithelium represents a huge surface area of approximately 100 m<sup>2</sup> that is lined by a single layer of columnar IECs, which form a stout physical barrier. IECs, however, form much more than a simple physical barrier that processes and absorbs dietary nutrients. They perform several other functions that are crucial for intestinal homeostasis<sup>3</sup>. These include secretion of compounds that influence microbial colonization, sampling of the intestinal microenvironment, sensing of both beneficial and harmful microbes, and induction and modulation of immune responses. To fulfil such diverse functions, the intestinal epithelium has unique anatomical and cellular adaptations, and IECs comprise several specialized cell types with distinct functions<sup>3,4</sup>. In addition, IECs do not regulate intestinal homeostasis in a solely intrinsic fashion, but instead function as part of a coordinated response to signals provided by the commensal microbiota and from local leukocyte populations<sup>3,4</sup>.

The mucosal surfaces of the body, including the intestine, are coated by a thick mucus gel comprising an outer layer of secreted mucins overlying a dense inner glycocalyx of membrane-anchored mucins that is inaccessible to most bacteria<sup>3,4</sup>. In addition to providing a biophysical barrier, mucus forms a matrix that allows the retention of high concentrations of antimicrobial molecules, such as defensins and secretory IgA, close to the epithelial surface. The mucus layer has a crucial role in intestinal homeostasis, as decreased levels of goblet cells, leading to reduced mucin secretion, are a hallmark of human IBD, and mice lacking the major mucin protein MUC2 develop spontaneous colitis<sup>1,3,4</sup>.

IECs also regulate the colonization and penetration of the epithelium by luminal microbes through the secretion of antimicrobial peptides (AMPs), which include lysozymes, defensins, cathelicidins, lipocalins and C-type lectins such as RegIIIγ<sup>5</sup>. Although some AMPs are produced constitutively by many IECs, others are secreted in an inducible fashion by Paneth cells — a type of specialized IEC located at the base of the intestinal crypts of the small intestine. In mice, Paneth cells have been implicated in protection against intestinal pathogens, as well as in limiting colonization by commensal bacteria<sup>6</sup>. Several lines of evidence suggest that Paneth cell dysfunction and impaired defensin secretion may contribute to IBD susceptibility. Patients with ileal Crohn's disease or those with *NOD2* (also known as *CARD15*)-susceptibility alleles had reduced α-defensin expression, and genetic variants in the transcription factor *TCF4*, which is involved in Paneth cell maturation and function, have recently been associated with ileal Crohn's disease<sup>5</sup>.

## Autophagy and ER stress control epithelial homeostasis

Paneth cell abnormalities have also been reported in patients with Crohn's disease who are homozygous for the T300A disease-risk allele of the autophagy gene *ATG16L1*, and in mice rendered hypomorphic for the *ATG16L1* protein<sup>7</sup>. Autophagy is a fundamental process that controls the catabolism of intracellular constituents in response to stress or infection, characteristically involving the formation of autophagosomes that target cargo for degradation by the lysosomal machinery<sup>8</sup>. Autophagy affects host immune responses on several levels and has a vital role in cell-intrinsic defence against intracellular infections<sup>8</sup>. Defects in *ATG16L1* led to the accumulation of morphologically abnormal granules in Paneth cells, suggesting

<sup>1</sup>Sir William Dunn School of Pathology, University of Oxford, South Parks Road, Oxford OX1 3RE, UK. <sup>2</sup>Translational Gastroenterology Unit, Experimental Medicine Division – NDM, University of Oxford, John Radcliffe Hospital, Headington, Oxford OX3 9DU, UK.

that the secretory granule pathway is impaired, although ATG16L1-deficient Paneth cells also expressed higher levels of pro-inflammatory mediators<sup>7</sup>.

Similarly, several recent studies have identified a link between endoplasmic reticulum (ER) stress and the consequent unfolded protein response (UPR) and IBD<sup>9</sup>. The maintenance of functional secretory cells requires coordinated protein folding and trafficking of secretory proteins by the ER–Golgi network. The UPR is elicited by the accumulation of unfolded or misfolded proteins in the ER, thus highly secretory cells, such as goblet cells and Paneth cells, are very susceptible to ER stress, and a functional UPR is required to maintain epithelial homeostasis in the gut<sup>9</sup>. Mice containing a genetic deletion in the IEC-restricted isoform of the UPR effector enzyme IRE1 showed increased susceptibility to colitis induced by dextran sulphate sodium (DSS) administration. Activation of IRE1 results in splicing and activation of the transcription factor XBP1, and mice with a conditional *Xbp1* deletion in IECs developed spontaneous enteritis that showed many characteristic features of human IBD<sup>10</sup>. Deletion of *Xbp1* in mouse IECs led to a loss of Paneth cells, a significant reduction in goblet cells and hyper-responsiveness of IECs to pro-inflammatory signals<sup>10</sup>. Other genetic lesions that result in increased ER stress in the intestinal epithelium also predispose individuals to intestinal inflammation<sup>9</sup>.

The ER stress pathway is also relevant to human gastrointestinal diseases, as increased ER stress is observed in the intestinal epithelium of patients with IBD, and polymorphisms within ER stress response genes, including *XBPI*, *AGR2* and *ORMDL3*, are associated with susceptibility to both Crohn's disease and ulcerative colitis<sup>9</sup>. The degree of ER stress within the intestinal epithelium may be modulated by environmental factors derived from the host or from the intestinal microbiota. Pro-inflammatory conditions and cytokines such as tumour necrosis factor- $\alpha$  (TNF- $\alpha$ ) exacerbate ER stress, whereas the anti-inflammatory cytokine interleukin-10 (IL-10) reduces it<sup>9</sup>. Furthermore, there is evidence to suggest that the ER acts as a source for the autophagosome membrane and that the UPR can activate autophagy<sup>9</sup>, indicating that these processes cooperate during infection or stress of the intestinal epithelium. Taken together, these studies indicate that defects in specialized secretory IEC populations, or aberrant ER stress or autophagy responses in IECs, greatly predispose people to the development of intestinal inflammation.

IECs may also influence intestinal homeostasis through the secretion of conditioning cytokines that affect adaptive responses primed by intestinal dendritic cells<sup>3</sup>. In the healthy intestine, these conditioning factors help to maintain a state of hyporesponsiveness towards commensal bacteria. For example, cytokines constitutively expressed by IECs, such as thymic stromal lymphopoietin and IL-25, limit dendritic-cell production of the p40 subunit of IL-12 and IL-23 and promote IL-10 secretion, impeding the priming of T helper 1 (T<sub>H</sub>1)-cell responses and instead favouring the induction of T regulatory (T<sub>reg</sub>)-cell and T<sub>H</sub>2-cell responses<sup>3,11</sup>. Conversely, after sensing pathogenic invasion or damage, IECs can elaborate the secretion of pro-inflammatory chemokines, such as IL-8 (also known as CXCL8), which have an important role in alerting the immune system to microbial attack<sup>3,12</sup>. IECs also exert a strong influence on local antibody responses by producing factors such as transforming growth factor- $\beta$  (TGF- $\beta$ ), B-cell activating factor (BAFF, also known as TNFSF13B) and a proliferation-inducing ligand (APRIL, also known as TNFSF13), which promote class-switching of B cells towards the production of IgA<sup>13</sup>. IECs mediate the transport of secretory IgA into the mucus layer, where it has a complementary role to innate defences in limiting the penetration of commensal bacteria across the epithelium<sup>13,14</sup>.

### Pattern recognition receptors and intestinal homeostasis

Accumulating evidence indicates that microbial sensing through pattern recognition receptors (PRRs) drives complementary functions in IECs and haematopoietic cells, which together control intestinal homeostasis<sup>12,15,16</sup>. The context of PRR activation is crucial. In the

healthy intestine, basal PRR activation maintains barrier function and commensal composition, but aberrant PRR signalling may be a central contributor to the pathophysiology of IBD. The latter point is emphasized by genetic-association studies linking PRR genes, including *NOD2*, *NLRP3* and various Toll-like receptor (TLR) genes, with IBD susceptibility, although the mechanisms responsible remain unclear<sup>12,16</sup>.

### Tonic PRR signals maintain a healthy epithelium

The importance of TLR signalling in regulating epithelial barrier function has been shown by studies using the DSS-colitis mouse model, in which DSS administration results in chemical destruction of the IEC layer and penetration of commensal bacteria, leading to acute colitis followed by restitution and repair of the epithelial barrier. Mice that lack specific TLRs, such as *Tlr2*, *Tlr4*, *Tlr5* or *Tlr9*, or that are deficient in the shared TLR signalling adaptor protein MyD88 show increased susceptibility to DSS colitis, characterized by defective tissue repair and/or increased mortality<sup>12,16</sup>. TLR signals drive intrinsic protective effects in IECs by inducing several proliferative and anti-apoptotic factors and by promoting epithelial restitution and fortifying intercellular tight junctions<sup>12,16</sup>. Intrinsic TLR signals in IECs also have a central role in limiting bacterial colonization and translocation by stimulating IEC production of AMPs, such as defensins and RegIII $\gamma$ <sup>6,15</sup>. An elegant study in which MyD88 expression was selectively limited to Paneth cells showed that these cells sense commensal bacteria directly through TLRs, and that this sensing induced AMP production that limited bacterial translocation across the intestinal mucosa<sup>6</sup>.

The activation of cytosolic NOD-like receptors (NLRs) may also be important in maintaining barrier function, as mice lacking *Nod1* or *Nod2*, or bearing a Crohn's-disease-associated mutant allele of human *NOD2*, showed defects in defensin secretion and increased susceptibility to DSS colitis<sup>17,18</sup>. Recent studies have indicated that NLR-mediated inflammasome activation also contributes to protection after damage of the epithelium, because mice deficient in *NLRP3*, its adaptor ASC or caspase-1 showed enhanced colitis and mortality after DSS administration<sup>19–21</sup>. Inflammasomes are multimolecular complexes that activate caspase-1. They are formed after the activation of cytosolic NLRs, which then associate with caspase-1, often through interactions with adaptor proteins such as ASC<sup>22</sup>. Activated caspase-1 has a central involvement in the processing and secretion of two key pro-inflammatory cytokines, IL-1 $\beta$  and IL-18, which in turn bind to receptors that use MyD88 as a signal-transduction adaptor<sup>22</sup>. The administration of exogenous IL-18 attenuated DSS colitis in *Casp1*<sup>−/−</sup> mice (which lack caspase-1)<sup>19,20</sup>, and *Il18*<sup>−/−</sup> mice and *Il18r1*<sup>−/−</sup> mice (which lack the IL-18 receptor) showed increased susceptibility to DSS colitis, highlighting a role for IL-18 in epithelial barrier integrity and repair<sup>23,24</sup>.

### Sustained PRR signals drive intestinal inflammation

In contrast to the protective barrier responses elicited by tonic PRR signals in IECs, studies in mouse models of chronic colitis have demonstrated that sustained PRR activation drives chronic intestinal inflammation. Thus, MyD88 signals were required for spontaneous colitis development in *Il10*<sup>−/−</sup> mice<sup>12,16</sup>. A key role for MyD88 signals in haematopoietic cells was indicated by experiments in which selective ablation of MyD88 rendered mice refractory to intestinal inflammation induced by *Helicobacter hepaticus*<sup>15</sup>. Although these results do not exclude a contribution from PRR signals in IECs to inflammatory responses in the gut, they indicate that such responses alone are insufficient to drive chronic inflammatory pathology. Similarly, recent studies have demonstrated that T-cell-intrinsic MyD88 signals are not essential for the expression of pathogenic effector or regulatory functions in the intestine<sup>15</sup>, whereas MyD88 signals in dendritic cells during the sensing of intestinal microbiota were shown to be essential for T-cell proliferation and intestinal pathology<sup>25</sup>.

Patients with IBD, particularly those with ulcerative colitis, have a greatly increased risk of developing colitis-associated cancer. Studies of the role of PRR signals in colitis-associated cancer, however, have



produced conflicting results<sup>26,27</sup>. Although sustained PRR-driven inflammatory responses can exacerbate intestinal tumorigenesis, their role in epithelial barrier maintenance and repair is protective against the development of intestinal tumours<sup>26,27</sup>. It is also worth noting that a few studies have reported protective effects of PRR signals against chronic intestinal inflammation, again emphasizing that some PRR signalling can be beneficial in the gut. For example, bacterial polysaccharide A derived from the human commensal bacterium *Bacteroides fragilis* was able to protect mice from T-cell-mediated colitis in a TLR2-dependent manner through the induction of T<sub>reg</sub> cells<sup>28</sup>.

### Integration of bacterial handling and stress responses

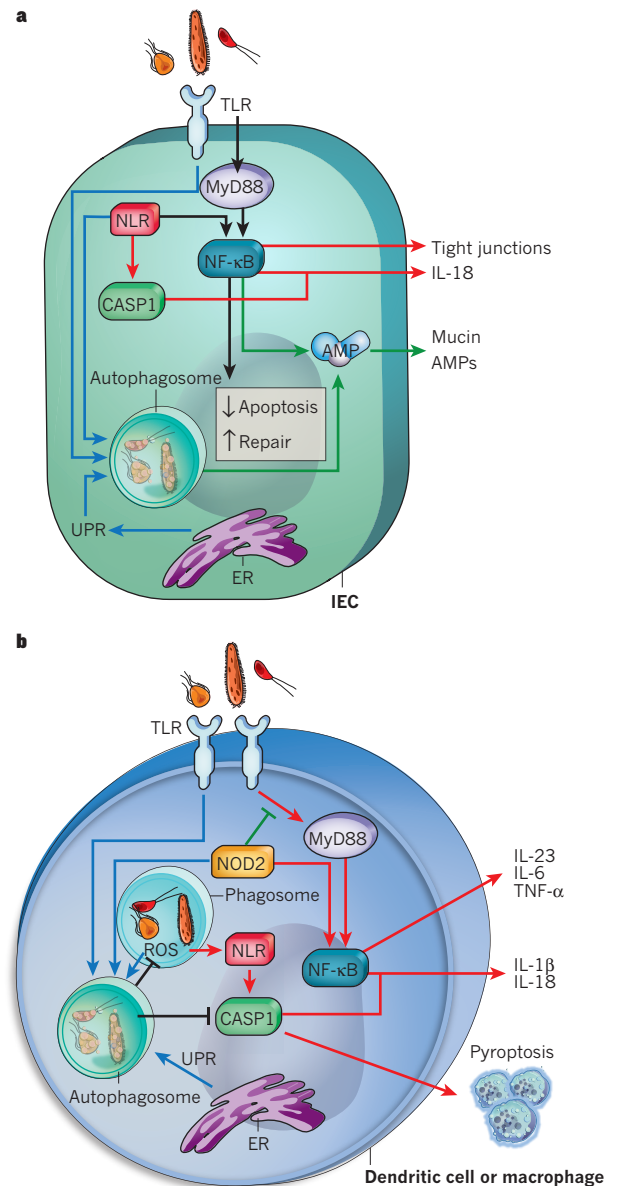
Microbes trigger a diverse range of PRRs and cellular stress responses that do not operate in isolation, but are integrated by the cell to direct appropriate effector responses. Thus, defects in one PRR pathway may influence other PRR signalling cascades, as well as affecting other processes implicated in intestinal homeostasis, such as autophagy and ER stress.

For example, Crohn's-disease-associated mutations in the *NOD2* gene are mainly located in the leucine-rich-repeat region that mediates sensing of the peptidoglycan motif muramyl dipeptide (MDP), resulting in reduced activation of nuclear factor- $\kappa$ B (NF- $\kappa$ B)<sup>17</sup>. It has been proposed that *NOD2* could act by attenuating TLR signalling that drives excessive activation of dendritic cells and T<sub>H</sub>1-cell responses<sup>17</sup>. Further evidence that *NOD2* can regulate TLR signalling in the gut came from a mouse model of necrotizing enterocolitis, in which activation of *NOD2* by MDP inhibited TLR4 signalling in IECs and ameliorated the condition<sup>29</sup>.

There is an increasing appreciation that PRR signals intersect with other bacterial-handling and cellular-stress processes to coordinate protective and inflammatory responses. For example, TLR signals can induce autophagy, and this has been reported to enhance the clearance of microbes<sup>8</sup>. Recent studies have found that *NOD2* stimulates autophagy by interacting directly with ATG16L1, which allows the recruitment of ATG16L1 to sites of bacterial entry<sup>30,31</sup>. Furthermore, dendritic cells expressing Crohn's-disease-associated mutant forms of *NOD2* or *ATG16L1* showed reduced autophagy in response to MDP, and this led to impaired antigen presentation and bacterial killing<sup>31</sup>.

The regulation of inflammasome activation by autophagic pathways has been reported, with selective ablation of ATG16L1 in mouse haematopoietic cells leading to increased inflammasome activation and IL-1 $\beta$  secretion in response to lipopolysaccharide<sup>32</sup>. These mice were also highly susceptible to DSS colitis, which was reversed by the neutralization of IL-1 and IL-18 (ref. 32). The mechanism involved is not clear, but autophagy has been reported to inhibit the generation of reactive oxygen species (ROS), especially by dysfunctional mitochondria, which have been shown to trigger the activation of NLRP3 inflammasomes<sup>33</sup>. Autophagy may also inhibit pyroptosis, a highly inflammatory form of caspase-1-dependent cell death that has been observed in myeloid cells infected with intracellular pathogens<sup>8</sup>. In addition, ROS generation has been shown to stimulate autophagy that restricted the replication of *Salmonella enterica* serovar Typhimurium (*S. Typhimurium*) in IECs, suggesting that ROS-induced autophagy may act as a negative-feedback mechanism to limit caspase-1-driven inflammatory circuits, while providing a complementary mechanism of bacterial defence<sup>22</sup>. Recent studies have also reported interactions between TLR and ER stress pathways, although both inhibition and activation of distinct arms of the UPR have been observed after TLR stimulation<sup>9,34</sup>.

Taken together, these studies indicate that diverse PRR signals interact with the autophagy and ER stress pathways to coordinate bacterial handling and inflammatory responses, and suggest that deficiencies or perturbations of these networks could contribute to IBD pathogenesis (Fig. 1). A better understanding of how these networks function in IECs and leukocytes in the healthy and inflamed intestine may give rise to new therapeutic avenues for IBD, and could also reveal strategies for boosting mucosal barrier defences in immune-suppressed individuals.



**Figure 1 | Bacterial sensing and cellular stress pathways in intestinal homeostasis.** **a**, Bacterial sensing and handling pathways cooperatively maintain the intestinal epithelial barrier. In IECs, basal sensing of microbial pathogen-associated molecular patterns by PRRs such as TLRs maintains epithelial barrier function by stimulating AMP expression (green arrows), and by fortifying tight junctions and inducing the release of protective cytokines such as IL-18 (red arrows). In addition, intrinsic PRR signals in IECs stimulate anti-apoptotic and proliferative responses (black arrows). The autophagy pathway, induced by PRR signals or by the ER stress response (blue arrows), cooperates with PRRs to promote the secretion of AMP and mucins (green arrows), and also constitutes an important cell-intrinsic defence mechanism in IECs for bacterial clearance. Therefore, defects in PRR sensing, the UPR or the autophagy pathways can result in impaired barrier function, leading to increased bacterial colonization and translocation and eliciting an exacerbated inflammatory response. **b**, Aberrant PRR signals in haematopoietic cells drive chronic inflammation in IBD. PRR signals in dendritic cells and macrophages drive chronic inflammatory responses in the gut through the activation of NF- $\kappa$ B-dependent pro-inflammatory cytokines (such as IL-23, IL-6 and TNF- $\alpha$ ) and caspase-1 (CASP1)-mediated induction of IL-1 $\beta$ , IL-18 and pyroptosis (red arrows). The magnitude of these inflammatory responses may be tempered by *NOD2* suppression of TLR signals (green line) and stimulation of the autophagy pathway by *NOD2*, TLRs or ROS induction (blue arrows). Autophagy may attenuate ROS production and CASP1 activation (black lines), thus limiting inflammatory responses. Therefore, defects in *NOD2* or autophagy pathways may contribute to the excessive inflammatory responses observed in IBD.

## Pathological effector modules in the gut

There are a multitude of animal models of IBD that either arise spontaneously or are induced by various experimental manipulations, which reproduce distinct features of human IBD. However, there is no perfect experimental model, because patients with IBD present a heterogeneous spectrum of pathological features that reflect the participation of a diverse range of innate and adaptive immune effectors. This heterogeneity is further underscored by the recent observations that around 100 distinct genetic loci may contribute to IBD susceptibility (see page 307), and the key target of these aberrant immune responses, the gut microbiota, is unique to each individual<sup>2</sup>. It is therefore likely that there will be several aetiologies of human IBD, and these may reflect aberrant expression of distinct immune modules. In this context, an immune module is used to define an effector response coordinated by a group of cytokines that may be produced by innate and/or adaptive leukocytes. These distinct immune modules evolved to protect against the different types of challenge posed by diverse pathogens that target the gastrointestinal tract, and some immune pathology associated with their expression is one 'cost' of a functional immune system (Fig. 2).

### TNF- $\alpha$ and IL-6

The central role of innate myeloid cells in IBD is mirrored by the potent pro-inflammatory effects of the cytokines that they secrete, particularly TNF- $\alpha$  and IL-6. The successful application of anti-TNF- $\alpha$  antibody treatment signified a major breakthrough in the treatment of IBD<sup>35</sup>, and resulted directly from convergent data in animal IBD models indicating a role for TNF- $\alpha$  in chronic intestinal inflammation<sup>1</sup>. IL-6 is increased in the inflamed intestinal mucosa, and blockade of IL-6 signalling ameliorated colitis in mouse models and also had beneficial effects in a clinical trial of patients with Crohn's disease<sup>1</sup>. Despite the success of anti-TNF- $\alpha$  biologics in IBD, approximately one-third of patients do not respond to anti-TNF- $\alpha$  treatment, and many others eventually lose responsiveness or become intolerant to these agents<sup>35</sup>. In addition, patients treated with anti-TNF- $\alpha$  show an increased incidence of severe infections and malignancies, emphasizing the need for further therapies that may target intestinal inflammation more selectively<sup>35</sup>.

### IL-23, T<sub>H</sub>1 and T<sub>H</sub>17 responses

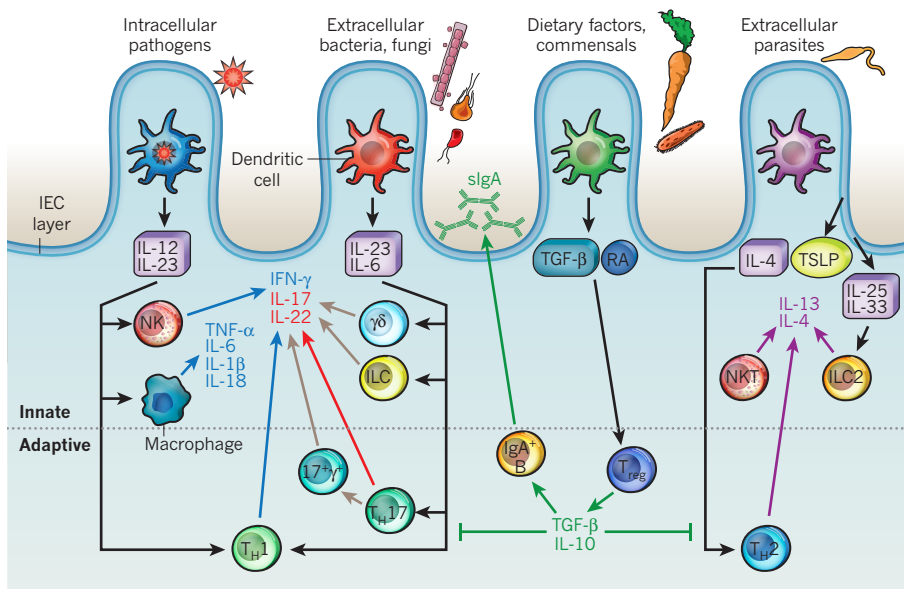
Early studies on the production of T-cell-derived cytokines suggested a role for IL-13-producing natural killer T (NKT) cells in ulcerative colitis, and the differential activation of IL-12 p40 and T<sub>H</sub>1-cell responses was associated with Crohn's disease<sup>1</sup>. The colitis-attenuating effects observed in mice lacking the *Il12b* (also known as *Il-12p40*) gene, or given neutralizing antibodies directed against IL-12 p40 or interferon- $\gamma$  (IFN- $\gamma$ ), also emphasized a key role for T<sub>H</sub>1-cell responses in intestinal inflammation<sup>1</sup>. However, the discovery that another IL-12-p40-containing heterodimeric cytokine, IL-23, was the central driver in several autoimmune pathologies prompted analysis of the role of IL-23 in intestinal inflammation. Studies in several mouse IBD models have used selective targeting of the IL-23 p19 subunit to demonstrate that IL-23 plays a key part in chronic intestinal pathology<sup>36</sup>. These findings were quickly followed by genome-wide association studies (GWAS) that reported strong associations of polymorphisms in the *IL23R* and *IL12B* gene loci with Crohn's disease and ulcerative colitis (see page 307). IL-23 is induced by PRR stimulation and is constitutively expressed in a small population of dendritic cells present in the lamina propria of the terminal ileum, although in patients with Crohn's disease, CD14<sup>+</sup> intestinal macrophages have also been reported to secrete large amounts of IL-23 (ref. 37). The factors that determine whether an activated dendritic cell will preferentially produce IL-23 or IL-12 are not clear, but a recent study showed that ER stress and activation of the UPR can synergize with TLR signals to selectively increase IL-23 expression by dendritic cells<sup>38</sup>. Although IL-23 was initially linked to the preferential expression of T<sub>H</sub>17 responses, it can promote a wide range of pathological responses in the intestine, mediated either by T cells or by excessive innate immune activation<sup>36,39</sup>. IL-23-mediated enhancement

of T<sub>H</sub>1 and T<sub>H</sub>17 responses is consistent with the increased levels of IFN- $\gamma$ , IL-17 and IL-22 observed in the chronically inflamed intestine<sup>36,39</sup>. Transcription factors that direct T<sub>H</sub>1-cell or T<sub>H</sub>17-cell responses — such as T-bet (also known as T-box protein 21 and TBX21) or retinoic-acid-receptor-related orphan receptor- $\gamma$ t (ROR $\gamma$ t), respectively — were shown to be essential for T-cell-mediated colitis<sup>1,40</sup>. T-cell-intrinsic IL-23R signals favour the expression of pathogenic pro-inflammatory T-cell responses in several ways, including enhanced proliferation of effector T cells, reduced differentiation of FOXP3<sup>+</sup> T<sub>reg</sub> cells and the emergence of IL-17<sup>+</sup>IFN- $\gamma$ <sup>+</sup>CD4<sup>+</sup> T cells<sup>39</sup>. The precise origins, stability and pathogenic properties of these IL-17<sup>+</sup>IFN- $\gamma$ <sup>+</sup>CD4<sup>+</sup> T cells remain to be determined. Notably, IL-17<sup>+</sup>IFN- $\gamma$ <sup>+</sup>CD4<sup>+</sup> T cells have been isolated from the inflamed lamina propria of patients with Crohn's disease, and were shown to respond to IL-23 and to be derived from a discrete subset of CD161<sup>+</sup>CD4<sup>+</sup> T cells that express chemokine receptor 6 (CCR6) and the nuclear receptor ROR $\gamma$ t<sup>41,42</sup>. In phase II clinical trials, anti-IL-12-p40 monoclonal antibodies have shown clinical efficacy in a subset of patients with Crohn's disease, particularly in those who had not responded to anti-TNF- $\alpha$  therapy, suggesting that TNF- $\alpha$  and IL-12 (or IL-23) drive distinct pathways of immune pathology<sup>35</sup>.

The relative enrichment of T<sub>H</sub>17 cells at mucosal sites, together with the increased levels of T<sub>H</sub>17 cytokines in the inflamed gut, has fuelled interest in their potential role in IBD pathogenesis<sup>36,39</sup>. T<sub>H</sub>17 cells produce several cytokines, including IL-17A, IL-17F, IL-21 and IL-22 (ref. 36). Many studies have focussed on the roles of IL-17A and IL-17F, which are known to have pro-inflammatory effects in tissues such as the lung and brain, through the elaboration of cytokines and chemokines, particularly those that promote neutrophil recruitment<sup>36</sup>. Analyses of IL-17A and IL-17F in mouse colitis models have produced conflicting results. In acute DSS colitis, IL-17A has a protective role, whereas IL-17F seems to exacerbate disease<sup>43</sup>. By contrast, the neutralization of IL-17A attenuated chronic colitis in mice with *Stat3*-deficient T<sub>reg</sub> cells<sup>44</sup> and decreased innate immune colitis after *H. hepaticus* infection<sup>45</sup>. Studies in T-cell-transfer colitis models suggest that IL-17A and IL-17F can have redundant pro-inflammatory effects in the gut<sup>40</sup>. T<sub>H</sub>17 responses were also recently implicated in a model of colitis-associated cancer, because IL-17A depletion reduced colitis and tumour development<sup>46</sup>. The microbiota has an important role in the preferential localization of T<sub>H</sub>17 cells in the gut, as the colonization of germ-free mice with segmented filamentous bacteria led to the marked accumulation of T<sub>H</sub>17 cells in the intestinal lamina propria<sup>47,48</sup>.

Another T<sub>H</sub>17 cytokine that is highly dependent on IL-23 is IL-22, which enhances the innate immunity of tissues. Expression of the IL-22R complex is restricted to non-haematopoietic cells, especially epithelial cells in the skin, gut and lungs<sup>49</sup>. IL-22 signalling in IECs drives the production of AMPs and also promotes epithelial regeneration and healing by activating the transcription factor STAT3 (ref. 50). Consistent with this epithelial-protective role, IL-22 administration attenuated disease severity in the DSS and T-cell receptor- $\alpha$  (*Tcr $\alpha$* <sup>-/-</sup>) mouse IBD models, by restoring goblet cells and mucus production<sup>49</sup>. By contrast, other studies support a pathogenic role for IL-22 in IBD, as its expression is increased in patients with Crohn's disease, and high serum IL-22 levels correlate with increased disease activity and susceptibility-associated *IL23R* polymorphisms<sup>49</sup>. A recent study of bleomycin-induced airway inflammation showed that IL-22 could mediate either tissue-protective or pathogenic functions, depending on the absence or presence of IL-17A, respectively<sup>51</sup>. Thus, further studies are required to determine whether IL-17A or other pro-inflammatory mediators can have similar modulating effects on IL-22 activity in the gut. Although less extensively studied, IL-21 may also regulate intestinal inflammation, through effects on T<sub>H</sub>17 cells and the production of matrix metalloproteinases, which are involved in tissue remodelling<sup>36</sup>. In summary, although it is clear that T<sub>H</sub>17 cytokines are important in many aspects of intestinal homeostasis and protection from mucosal pathogens, their role in IBD pathogenesis remains ambiguous, and further investigations are necessary to clarify their potential for therapeutic intervention.





**Figure 2 | Conserved innate and adaptive immune effector modules in the gut.** IECs and intestinal dendritic cells sense distinct infectious agents, leading to the production of factors that direct different effector responses (black arrows). Local innate leukocyte populations can rapidly produce these effector cytokines to restrict pathogen growth until specific adaptive responses have been induced. Cells and molecules associated with distinct effector profiles are colour-coded; T<sub>H</sub>1 type (blue), T<sub>H</sub>17 type (red), mixed T<sub>H</sub>1 and T<sub>H</sub>17 (brown), T<sub>H</sub>2 type (purple), and aberrant or prolonged expression of any of these modules may contribute to chronic intestinal inflammation. Regulatory T-cell circuits (green) can suppress all types of inflammatory effector response and may enhance the production of protective secretory IgA (sIgA) antibodies. 17<sup>γ</sup>, IL-17- and IFN-γ-secreting CD4<sup>+</sup> T cell; ILC2, type-2 ILC; NK, natural killer; RA, retinoic acid; TSLP, thymic stromal lymphopoietin.

### NLRs and inflammasome-associated cytokines

Recent advances identifying a central role for inflammasomes and NLRs in autoinflammatory diseases<sup>22</sup> — together with the association of IBD with polymorphisms in *NLRP3* and *IL18RAP* — have rekindled interest in the potential roles of IL-1β and IL-18 in IBD. Levels of IL-1β and IL-18 are increased in IBD<sup>1,23</sup>, and *Il18*<sup>-/-</sup> mice were resistant to colitis induced by trinitrobenzene sulphonate<sup>24</sup>, suggesting that IL-1β and IL-18 enhance chronic intestinal pathology. This hypothesis is consistent with the ability of IL-1β and IL-18 to promote T<sub>H</sub>17 and T<sub>H</sub>1 responses, respectively<sup>23,52</sup>, and with studies indicating that ATG16L1 and α-defensins negatively regulate IL-1β expression<sup>32</sup>. Furthermore, this inflammatory axis is important in responses to gut pathogens, as ASC-mediated IL-1β production has an essential role in *Clostridium difficile* toxin-induced intestinal pathology<sup>53</sup>, and both IL-1β and IL-18 were required for the induction of intestinal inflammation after infection with *S. Typhimurium*<sup>54</sup>. Thus, inflammasome-forming NLRs can contribute to intestinal pathology through IL-1β and IL-18, and further studies are required to define their roles in IBD<sup>23</sup>.

### Conserved innate and adaptive effector modules

Although predominantly attributed to CD4<sup>+</sup> T<sub>H</sub> cells, many innate leukocyte populations, including γδ T cells, NKT cells and natural killer cells, can secrete T<sub>H</sub>1 and T<sub>H</sub>17 cytokines such as IFN-γ, IL-17A and IL-22 (refs 36, 49, 55–57). In particular, γδ T cells can express the T<sub>H</sub>17-associated transcription factors RORγt and the aryl hydrocarbon receptor (AHR), as well as the homing receptor CCR6, and can secrete IL-17 and IL-22 in response to IL-23 and IL-1β<sup>58,59</sup>.

However, recent studies have converged on the identification of several new innate lymphoid cell (ILC) populations present in the gut that can produce these pro-inflammatory cytokines<sup>55,56,60</sup>. Although many functionally heterogeneous ILC populations have been described, their phenotypic characteristics suggest that they are related to natural killer cells and lymphoid tissue inducer (LTI) cells<sup>55,56,60</sup>. Indeed, a recent cell-fate-mapping study suggested that several natural-killer-like and LTI-like ILC subsets were derived from a common RORγt<sup>+</sup> precursor<sup>61</sup>, and functional specialization is coordinated by cytokines and microbiota-dependent signals that direct the expression of distinct transcription factors<sup>55,56,60</sup>. In terms of IBD, we identified a population of CD90<sup>+</sup> CD4<sup>+</sup> ILCs that accumulated in the inflamed colons of *Rag*<sup>-/-</sup> mice infected with *H. hepaticus*<sup>45</sup>. These cells expressed high levels of IL-23R and RORγt and produced IFN-γ, IL-17A and IL-22 in response to IL-23, and their depletion with an anti-CD90 monoclonal antibody led to the attenuation of typhlocolitis<sup>45</sup>. A similar population of IL-23-responsive CD90<sup>+</sup> CD4<sup>+</sup> LTI cells was recently shown to be

an essential source of protective innate IL-22 during the initial phase of *Citrobacter rodentium* infection<sup>62</sup>. Other studies have identified further populations of innate leukocytes in the gut and mesenteric adipose tissues that secrete large amounts of the T<sub>H</sub>2 signature cytokines IL-4 and IL-13 (refs 63–65). These type-2 ILCs are RORγt<sup>-</sup> and were rapidly activated in response to IEC-derived IL-25 and IL-33 after infection with parasitic helminth worms, an infection in which T<sub>H</sub>2 cytokines are key contributors to protective immunity<sup>64,65</sup>. Although the precise relationships of these cells to one another, and to the RORγt<sup>+</sup> ILC populations discussed earlier, remain to be determined, these results further emphasize that various populations of intestinal innate leukocytes can rapidly respond to different types of infection by producing appropriate effector cytokines.

The emerging data on shared innate and adaptive cytokine profiles suggest that these conserved immune modules evolved before adaptive lymphocytes, and that protective immunity can be mediated by sentinel tissue-resident innate leukocytes early after infection, whereas subsequent T-cell responses add memory and specificity to the relevant protective axis (Fig. 2).

### Adaptive regulation of intestinal inflammation

The intestine contains an extensive network of dendritic cells and macrophages that has an important role in shaping adaptive immunity in response to intestinal environmental cues<sup>66,67</sup>. Under homeostatic conditions, both dendritic cells and macrophage populations have specific adaptations that promote tolerance. During infection, however, responses shift to a more inflammatory nature, which can lead to immune pathology when dysregulated.

### Antigen-presenting myeloid cells

Intestinal myeloid antigen-presenting cell (APC) populations are heterogeneous in terms of phenotype, function, developmental origin and anatomical location<sup>66,67</sup>. Recently, two major populations of intestinal dendritic cells have been identified on the basis of differential expression of the integrin subunit CD103 and the chemokine receptor CX<sub>3</sub>CR1. CD11c<sup>high</sup> CD103<sup>+</sup> dendritic cells share developmental origins with lymphoid tissue dendritic cells and are derived from pre-dendritic cells without a monocyte intermediate<sup>68,69</sup>. By contrast, monocytes give rise to intestinal CD11c<sup>high</sup> CD103<sup>-</sup> CX<sub>3</sub>CR1<sup>+</sup> dendritic cells, suggesting a close relationship between these cells and CX<sub>3</sub>CR1<sup>+</sup> intestinal macrophages. CD103<sup>+</sup> dendritic cells are dispersed throughout the lamina propria and in organized lymphoid structures. In the small intestine, they act as important sentinel cells as they can take up pathogenic and commensal bacteria, as well as innocuous antigens or

apoptotic IECs. After maturation, CD103<sup>+</sup> dendritic cells migrate to the draining mesenteric lymph node (MLN), where they initiate adaptive responses<sup>69,70</sup> focused on the intestine, including upregulation of the gut-homing receptors CCR9 and  $\alpha_4\beta_7$ -integrin on activated T cells, and IgA class-switch recombination by intestinal B cells<sup>66,67</sup>. These properties depend on the production of the vitamin A dietary metabolite retinoic acid, and CD103<sup>+</sup> dendritic cells express the retinal-metabolizing enzyme genes *ALDH1A1* and *ALDH1A2*, although it is not known whether CD103<sup>+</sup> dendritic cells are an essential functional source of retinoic acid *in vivo*<sup>66</sup>.

CD103<sup>+</sup> dendritic cells preferentially induce tolerance pathways, including FOXP3<sup>+</sup> T<sub>reg</sub> cells in the draining MLN by a TGF- $\beta$ -dependent and retinoic-acid-dependent mechanism<sup>66</sup>. The intestinal pathways that influence CD103<sup>+</sup> dendritic cell function *in vivo* are poorly understood, but they do not seem to depend on the microbiota<sup>71</sup>. A recent study showed that ablation of  $\beta$ -catenin expression in mouse dendritic cells led to reduced frequencies of T<sub>reg</sub> cells and higher frequencies of effector T<sub>H</sub>1 and T<sub>H</sub>17 cells in the intestinal lamina propria<sup>72</sup>. This correlated with reduced messenger RNA levels of *Il10*, *Tgfb*, *Aldh1a1* and *Aldh1a2*, resulting in enhanced susceptibility to DSS colitis. How  $\beta$ -catenin expression is regulated in intestinal dendritic cells is not known.

The functional properties of CD103<sup>+</sup> dendritic cells are not hardwired, and they acquire inflammatory properties during intestinal inflammation such as the ability to produce IL-6 and drive T<sub>H</sub>1 responses<sup>73</sup>. Thus, migratory CD103<sup>+</sup> dendritic cells can promote both tolerogenic and effector T-cell responses, and further work is required to identify quantitative and qualitative factors that drive intestinal dendritic cell conditioning and the effect of these factors on adaptive effector pathways.

CX<sub>3</sub>CR1<sup>+</sup>CD103<sup>+</sup> APCs comprise a heterogeneous population of dendritic cells and macrophages. CD11c<sup>+</sup>CX<sub>3</sub>CR1<sup>+</sup> dendritic cells are present adjacent to the intestinal epithelium and can extend processes through the epithelium to sample antigens and bacteria<sup>66,67</sup>. However, CD11c<sup>+</sup>CX<sub>3</sub>CR1<sup>+</sup> dendritic cells do not seem to migrate to the MLN and fail to prime naive T cells, suggesting that their main role may be to modulate local adaptive intestinal responses<sup>70</sup>. CX<sub>3</sub>CR1<sup>+</sup> APCs accumulate in response to microbiota-derived signals<sup>71</sup>, and a CD70<sup>high</sup> subset promotes colonic T<sub>H</sub>17 responses in response to commensal-derived ATP<sup>67</sup>. Colonic macrophages contribute to intestinal homeostasis in several ways. As highly phagocytic cells, they clear apoptotic cells and debris and contribute to wound repair of the epithelium<sup>74</sup>. Intestinal macrophages have adaptations to prevent excessive inflammatory responses towards the intestinal flora, including expression of inhibitors of NF- $\kappa$ B signalling that permit bactericidal activity in the absence of TLR-driven pro-inflammatory cytokine production<sup>74</sup>. Recent evidence suggests that these cells also promote tolerance in part through IL-10 production and maintenance of FOXP3 among colonic T<sub>reg</sub> cells<sup>75</sup>. A similar APC population in the small intestine induced FOXP3<sup>+</sup> T<sub>reg</sub> cells *in vitro*<sup>66</sup> and promoted T<sub>reg</sub>-cell proliferation in a CX<sub>3</sub>CR1-dependent manner<sup>76</sup>.

In IBD and experimental colitis, there is an increase in dendritic cell and macrophage populations that may contribute to intestinal pathology through pro-inflammatory cytokine production<sup>67</sup>. Although it is not yet fully established whether this reflects changes in resident myeloid cell populations or the accumulation of newly recruited cells, there is evidence to support the latter. Thus, acute and chronic mouse colitis models were associated with a marked increase in recruited monocyte-derived dendritic cells that produced IL-12, IL-23 and TNF- $\alpha$  and showed enhanced TLR responsiveness<sup>68,77</sup>. A similar population of inflammatory macrophages that promote colonic inflammation has also been described<sup>78</sup>. The data are consistent with a model in which sustained pro-inflammatory cytokines and chemokines promote myelopoiesis — the mobilization of monocytes from the bone marrow to the blood and the recruitment of inflammatory monocytes to the inflamed intestine. These recruited myeloid cells lack gut-specific adaptations associated with tolerance and instead mediate inflammatory

responses after microbial challenge. Further understanding of the factors that control the recruitment and function of myeloid cells in intestinal inflammation may provide new therapeutic targets.

### Regulatory T-cell populations

Although various T-cell populations have anti-inflammatory functions, FOXP3<sup>+</sup> T<sub>reg</sub> cells and FOXP3<sup>+</sup> IL-10-secreting CD4<sup>+</sup> T cells are particularly important in the intestine<sup>79</sup>. Most of the former acquire FOXP3 expression in the thymus and represent a functionally distinct population that has a non-redundant role in controlling immune homeostasis. Deletion or loss-of-function mutations in the gene encoding FOXP3 result in a fatal inflammatory disease in mice, and in immune dysregulation, polyendocrinopathy, enteropathy, X-linked (IPEX) syndrome in humans, which is often accompanied by intestinal inflammation<sup>79</sup>. FOXP3<sup>+</sup> T<sub>reg</sub> cells are abundant in the small intestine and colon, where they control potentially deleterious responses to dietary and microbial stimuli<sup>79</sup>. In addition to thymic-derived T<sub>reg</sub> cells, the intestine is also a preferential site for TGF- $\beta$ -dependent induction of FOXP3<sup>+</sup> T<sub>reg</sub> cells from naive CD4<sup>+</sup> T-cell precursors<sup>79</sup>. Such antigen-induced T<sub>reg</sub> cells are expanded in models of oral tolerance and can control local and systemic antigen-induced hypersensitivity responses. Little is known about the antigen specificity of intestinal T<sub>reg</sub> cells that control microbiota-driven responses. However, T<sub>reg</sub>-cell accumulation in the colon is reduced in germ-free mice and can be increased by particular indigenous bacteria, suggesting a role for the microbiota in promoting intestinal T<sub>reg</sub>-cell responses<sup>80</sup>.

Induced T<sub>reg</sub>-cell and T<sub>H</sub>17-cell populations seem to be reciprocally regulated in the intestine. Although TGF- $\beta$  is required for the differentiation of both populations, the presence of STAT3-mediated signals (such as IL-6 or IL-23) promotes T<sub>H</sub>17 cells at the expense of FOXP3<sup>+</sup> T<sub>reg</sub> cells<sup>39,81</sup>. Such a mechanism allows the inflammatory response to override T<sub>reg</sub>-cell induction in the presence of pro-inflammatory stimuli, promoting intestinal effector T-cell responses and host defence. Recent evidence suggests that bacterial components differentially affect this balance, providing potential therapeutic strategies to influence tolerance and immunity in the gut<sup>28,82</sup>.

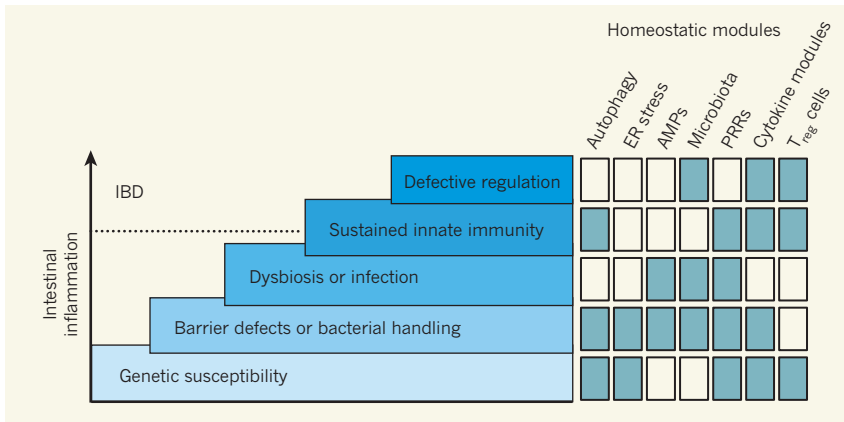
An important component of T<sub>reg</sub>-cell-mediated control of intestinal homeostasis is their ability to survive and compete with effector T cells in the intestinal niche<sup>81</sup>. This has recently been shown to involve the expression of co-stimulatory pathways and transcription factor modules associated with the colitogenic response<sup>44,83</sup>. For example, mice with a *Stat3* deletion in FOXP3<sup>+</sup> T<sub>reg</sub> cells develop aggressive colitis owing to uncontrolled T<sub>H</sub>17 responses<sup>44</sup>. In addition to STAT3, T<sub>reg</sub> cells can express several transcription factors associated with particular effector responses, including T-bet, IRF4 and GATA3 (ref. 81). Under homeostatic conditions these allow T<sub>reg</sub>-cell-mediated control of distinct effector modules. However, the system is delicately poised and can sometimes lead to T<sub>reg</sub>-cell instability. For example, high-level T-bet expression in the presence of acute intestinal infection drives T<sub>reg</sub> cells into an inflammatory IFN- $\gamma$ -secreting phenotype<sup>84</sup>.

### The immune-suppressive modules TGF- $\beta$ and IL-10

TGF- $\beta$  is present at high concentrations in the intestine and has a crucial involvement in modulating the immune response<sup>85</sup>. Deletion of *Tgfb1* in mice leads to a fatal inflammatory disease similar to FOXP3 deficiency. Cell-type-specific targeting of TGF- $\beta$  or its receptor has shown a key role for T cells in both the production and responsiveness to TGF- $\beta$  that is required to maintain immune homeostasis<sup>85</sup>. T cells that cannot respond to TGF- $\beta$  escape T<sub>reg</sub>-cell-mediated control, and T cells from patients with IBD are refractory to the anti-inflammatory actions of TGF- $\beta$  through expression of the negative regulator of TGF- $\beta$  signalling SMAD7 (ref. 86). Whether this is a primary or secondary event is not known, but it suggests that restoring TGF- $\beta$  responsiveness may have therapeutic benefit in IBD.

TGF- $\beta$  is produced as an inactive precursor that needs to be post-translationally modified to become biologically active. This is a





**Figure 3 | A multihit model of IBD pathogenesis.** The induction and perpetuation of chronic intestinal pathology may require the convergence of many abnormalities that affect several overlapping layers of immune homeostasis in the intestine. Defects in one layer are unlikely to precipitate IBD in the absence of further pathogenic lesions (panels on left). These layers are in turn controlled by a range of cell-intrinsic and cell-extrinsic regulatory modules that function in an interrelated way to maintain homeostasis (panels on right). Defects in these homeostatic modules may predispose people to the development of IBD by affecting several layers of immune homeostasis (filled panels). The dashed line denotes the threshold at which the level of inflammation manifests as IBD.

tightly controlled process that has recently been shown to involve expression of the  $\alpha_v\beta_8$ -integrin molecule on intestinal dendritic cells and macrophages<sup>66</sup>, as well as expression of the proprotein convertase furin on T cells<sup>87</sup>.  $\alpha_v\beta_8$ -Integrin has also been implicated in myeloid-cell uptake of apoptotic cells, a process that has been linked to TGF- $\beta$  production<sup>88</sup>. Because IECs undergo apoptosis under physiological conditions, this may provide a source of TGF- $\beta$  that promotes tolerance under homeostatic conditions. Enhanced apoptosis of IECs accompanies infection and inflammation, and under these circumstances TGF- $\beta$  in the presence of pro-inflammatory cytokines such as IL-6 promotes the development of inflammatory T<sub>H</sub>17 responses<sup>89</sup>. Lastly, a recent report suggests that TGF- $\beta$ -dependent stimulation of intestinal IgA responses is another mechanism through which T<sub>reg</sub> cells can reinforce intestinal homeostasis<sup>90</sup>.

IL-10 is produced by a wide range of leukocytes, including T cells, B cells and myeloid cells, and its deletion in mice leads to colitis development<sup>79</sup>. CD4<sup>+</sup> T-cell-produced IL-10 is required to prevent intestinal inflammation, with functional contributions from both FOXP3<sup>+</sup> and FOXP3<sup>-</sup> CD4<sup>+</sup> cells<sup>79</sup>. The intestine contains large numbers of CD4<sup>+</sup> IL-10<sup>+</sup> cells; in the colon, these are mainly FOXP3<sup>+</sup>, whereas both FOXP3<sup>+</sup> and FOXP3<sup>-</sup> IL-10<sup>+</sup> cells are present in the small intestine<sup>79</sup>. Intestinal bacteria can promote the activity of colonic T<sub>reg</sub> cells by inducing IL-10 production, and recent evidence suggests a specific role for particular *Clostridium* species in this process<sup>80</sup>. Unlike FOXP3-expressing cells, FOXP3<sup>-</sup> IL-10<sup>+</sup> CD4<sup>+</sup> cells may represent a more heterogeneous mix, because most effector T<sub>H</sub>-cell subsets, including T<sub>H</sub>1, T<sub>H</sub>2 and T<sub>H</sub>17 cells, produce IL-10 after chronic immune stimulation<sup>91</sup>. Myeloid sources of IL-10 are important in some settings, as IL-10 production by intestinal macrophages promoted FOXP3 T<sub>reg</sub>-cell function in an adoptive transfer model of colitis<sup>75</sup>. IL-10 controls chronic intestinal inflammation partly through direct anti-inflammatory effects on myeloid cells<sup>79</sup>. Evidence for the role of IL-10 in human IBD comes from findings that mutations in the IL-10 receptor genes *IL10RA* and *IL10RB* lead to severe early-onset IBD<sup>92</sup>. GWAS have also identified single nucleotide polymorphisms in *IL10* associated with susceptibility to Crohn's disease and ulcerative colitis (see page 307). Together, both genetic and functional studies highlight the importance of IL-10 in intestinal homeostasis and suggest that the ability of intestinal bacteria to induce IL-10 may be an important facet of host-commensal mutualism.

### A multihit model of IBD

As noted above, in some cases single gene defects in crucial regulatory circuits, such as the IL-10 pathway<sup>92</sup>, can trigger severe IBD in infants. However, the heterogeneity of IBD and the low disease penetrance in individuals carrying disease-susceptibility alleles suggest that, in most patients, several host and environmental factors interact to cause IBD. There is evidence to suggest that Crohn's disease stems from an immunodeficiency of macrophages that results in defective acute inflammatory responses and impaired clearance of commensal

bacteria, leading to the subsequent expression of chronic granulomatous inflammation<sup>93</sup>. Pathogenic infections may act as triggers or contributing factors for IBD, and adherent-invasive *Escherichia coli* are frequently present in close association with the ileal mucosa in patients with Crohn's disease<sup>1</sup>.

It is not yet clear whether the presence of *E. coli* is a cause or effect of colitis, as recent studies have highlighted that intestinal inflammation can confer a selective growth advantage to certain pathogens, including *S. Typhimurium*<sup>94</sup>. A growing body of evidence suggests that IBD is associated with an imbalance in the composition of the intestinal bacterial microbiota, termed dysbiosis<sup>2,95</sup>. Patients with IBD, particularly those with Crohn's disease, have alterations in the gut microbiota, with reduced diversity in major phyla, such as Firmicutes and Bacteroidetes, and increased numbers of Enterobacteriaceae<sup>2,95</sup>. A key unresolved issue is whether dysbiosis represents a primary or secondary predisposing factor for IBD, as it may be related to, or compounded by, other defects. Recent studies have indicated that dysbiosis is influenced both by the host genotype, such as the presence of *NOD2*- or *ATG16L1*-susceptibility alleles<sup>96</sup>, and by IBD phenotype, with patients with ileal Crohn's disease showing the most pronounced changes<sup>97</sup>. It is interesting that core commensals belonging to the Clostridiales order, such as *Faecalibacterium* and *Roseburia*, were significantly reduced in patients with ileal Crohn's disease<sup>96,97</sup>. These genera are potent sources of short-chain fatty acids, such as butyrate, that have been shown to have protective effects in mouse colitis models<sup>98</sup>. In addition, clostridial groups IV (which includes *Faecalibacterium*) and XIVa were recently shown to promote the accumulation of FOXP3<sup>+</sup> T<sub>reg</sub> cells in the mouse colon<sup>80</sup>. Dietary factors may also affect microbiota composition, leading to alterations in intestinal immune homeostasis<sup>98</sup>.

Taken together, these studies are beginning to illuminate some of the complex interactions between different host genetic and environmental factors that can predispose patients to the development of IBD. The induction and perpetuation of chronic intestinal pathology may require additive lesions that affect several layers of immune regulation in the gut (Fig. 3). Early animal model studies showed that an interaction between the intestinal flora and host factors is required for the development of intestinal inflammation. For example, colitis in IL-10-deficient mice requires the presence of triggering bacteria such as *H. hepaticus* infection<sup>79</sup>. Two recent studies further illustrate how multiple lesions may interact to elicit intestinal pathology. The first of these found that the Paneth cell abnormalities present in mice carrying a hypomorphic mutant of the *Atg16l1* autophagy gene (*Atg16l1*<sup>TM</sup>) were triggered by persistent infection with an enteric mouse norovirus strain (MNV CR6)<sup>99</sup>. In addition, *Atg16l1*<sup>TM</sup> mice infected with MNV CR6 developed exacerbated pathology after the administration of DSS, with characteristics of human Crohn's disease, including blunting of villi in the ileum, that depended on IFN- $\gamma$ , TNF- $\alpha$  and the presence of commensal bacteria<sup>99</sup>. Thus, Crohn's-disease-like pathology required persistent viral infection of a genetically susceptible host together with environmental factors and commensal bacteria. The second study

examined a model of transmissible ulcerative-colitis-like disease that arises in *T-bet*<sup>-/-</sup>*Rag2*<sup>-/-</sup> (also known as *Tbx21*<sup>-/-</sup>*Rag2*<sup>-/-</sup>) mice and is associated with defective colonic barrier function and hyperactivation of inflammatory dendritic cells<sup>100</sup>. Transmission of a milder degree of colitis to co-housed wild-type mice correlated with the presence of the Enterobacteriaceae species *Proteus mirabilis* and *Klebsiella pneumoniae*, but also required the presence of an endogenous microbiota<sup>100</sup>. Thus, maximal colitis involved barrier defects, hyperactivated innate immunity, an absence of T<sub>reg</sub> cells and alterations in the microbiota composition.

## Perspectives

Recent advances in mapping the genetic basis of disease susceptibility, coupled with rapid improvements in characterization of the microbiota in healthy and diseased individuals, offer great hope for the continued development of new IBD treatments. However, several key issues need to be understood better. These include distinguishing between individuality in IBD aetiology and commonality in pathogenic effector modules, so that therapies may be tailored to appropriate patient subgroups, such that distinct responses may be either suppressed or enhanced to restore homeostasis. The influence of microbiota-derived molecules on local and systemic immune responses is an area of great promise, but it will also be important to determine how immune responses feed back into shaping the composition of the microbiota and how different members of the microbiota interact within different environments in the gut, as well as to determine how to stably manipulate the gut microbiota. The real extent of effector T-cell plasticity *in vivo* and whether innate effector responses are similarly malleable needs to be investigated further, to establish whether the stable conversion of deleterious responses into beneficial ones may be achieved. Accomplishing these goals will require the cooperation of scientists working across several disciplines, an improved characterization of the pathophysiology of disease models and application of new technical approaches to clinical samples from patients with IBD. ■

- Kaser, A., Zeissig, S. & Blumberg, R. S. Inflammatory bowel disease. *Annu. Rev. Immunol.* **28**, 573–621 (2010).
- Hill, D. A. & Artis, D. Intestinal bacteria and the regulation of immune cell homeostasis. *Annu. Rev. Immunol.* **28**, 623–667 (2010).
- Artis, D. Epithelial-cell recognition of commensal bacteria and maintenance of immune homeostasis in the gut. *Nature Rev. Immunol.* **8**, 411–420 (2008).
- Hooper, L. V. & Macpherson, A. J. Immune adaptations that maintain homeostasis with the intestinal microbiota. *Nature Rev. Immunol.* **10**, 159–169 (2010).
- Kosloski, M. J., Beisner, J., Stange, E. F. & Wehkamp, J. Innate antimicrobial host defense in small intestinal Crohn's disease. *Int. J. Med. Microbiol.* **300**, 34–40 (2010).
- Vaishnava, S., Behrendt, C. L., Ismail, A. S., Eckmann, L. & Hooper, L. V. Paneth cells directly sense gut commensals and maintain homeostasis at the intestinal host-microbial interface. *Proc. Natl Acad. Sci. USA* **105**, 20858–20863 (2008).
- Cadwell, K. et al. A key role for autophagy and the autophagy gene *Atg16l1* in mouse and human intestinal Paneth cells. *Nature* **456**, 259–263 (2008).
- Deretis, V. & Levine, B. Autophagy, immunity, and microbial adaptations. *Cell Host Microbe* **5**, 527–549 (2009).
- Kaser, A., Martinez-Naves, E. & Blumberg, R. S. Endoplasmic reticulum stress: implications for inflammatory bowel disease pathogenesis. *Curr. Opin. Gastroenterol.* **26**, 318–326 (2010).
- Kaser, A. et al. XBP1 links ER stress to intestinal inflammation and confers genetic risk for human inflammatory bowel disease. *Cell* **134**, 743–756 (2008). **This report links ER stress in IECs to the development of intestinal inflammation in both mice and humans.**
- Taylor, B. C. et al. TSLP regulates intestinal immunity and inflammation in mouse models of helminth infection and colitis. *J. Exp. Med.* **206**, 655–667 (2009).
- Abreu, M. T. Toll-like receptor signalling in the intestinal epithelium: how bacterial recognition shapes intestinal function. *Nature Rev. Immunol.* **10**, 131–144 (2010).
- Fagarasan, S., Kawamoto, S., Kanagawa, O. & Suzuki, K. Adaptive immune regulation in the gut: T cell-dependent and T cell-independent IgA synthesis. *Annu. Rev. Immunol.* **28**, 243–273 (2010).
- Slack, E. et al. Innate and adaptive immunity cooperate flexibly to maintain host-microbiota mutualism. *Science* **325**, 617–620 (2009).
- Asquith, M. J., Boulard, O., Powrie, F. & Maloy, K. J. Pathogenic and protective roles of MyD88 in leukocytes and epithelial cells in mouse models of inflammatory bowel disease. *Gastroenterology* **139**, 519–529 (2010).
- Cario, E. Toll-like receptors in inflammatory bowel diseases: a decade later.

- Inflamm. Bowel Dis.* **16**, 1583–1597 (2010).
- Strober, W., Murray, P. J., Kitani, A. & Watanabe, T. Signalling pathways and molecular interactions of NOD1 and NOD2. *Nature Rev. Immunol.* **6**, 9–20 (2006).
- Chen, G. Y., Shaw, M. H., Redondo, G. & Nunez, G. The innate immune receptor Nod1 protects the intestine from inflammation-induced tumorigenesis. *Cancer Res.* **68**, 10060–10067 (2008).
- Dupaul-Chicoine, J. et al. Control of intestinal homeostasis, colitis, and colitis-associated colorectal cancer by the inflammatory caspases. *Immunity* **32**, 367–378 (2010).
- Zaki, M. H. et al. The NLRP3 inflammasome protects against loss of epithelial integrity and mortality during experimental colitis. *Immunity* **32**, 379–391 (2010).
- Allen, I. C. et al. The NLRP3 inflammasome functions as a negative regulator of tumorigenesis during colitis-associated cancer. *J. Exp. Med.* **207**, 1045–1056 (2010).
- Schroder, K. & Tschopp, J. The inflammasomes. *Cell* **140**, 821–832 (2010).
- Siegmund, B. Interleukin-18 in intestinal inflammation: friend and foe? *Immunity* **32**, 300–302 (2010).
- Salcedo, R. et al. MyD88-mediated signaling prevents development of adenocarcinomas of the colon: role of interleukin 18. *J. Exp. Med.* **207**, 1625–1636 (2010).
- Feng, T., Wang, L., Schoeb, T. R., Elson, C. O. & Cong, Y. Microbiota innate stimulation is a prerequisite for T cell spontaneous proliferation and induction of experimental colitis. *J. Exp. Med.* **207**, 1321–1332 (2010).
- Asquith, M. & Powrie, F. An innately dangerous balancing act: intestinal homeostasis, inflammation, and colitis-associated cancer. *J. Exp. Med.* **207**, 1573–1577 (2010).
- Saleh, M. & Trinchieri, G. Innate immune mechanisms of colitis and colitis-associated colorectal cancer. *Nature Rev. Immunol.* **11**, 9–20 (2011).
- Round, J. L. & Mazmanian, S. K. Inducible Foxp3<sup>+</sup> regulatory T-cell development by a commensal bacterium of the intestinal microbiota. *Proc. Natl Acad. Sci. USA* **107**, 12204–12209 (2010).
- Richardson, W. M. et al. Nucleotide-binding oligomerization domain-2 inhibits Toll-like receptor-4 signaling in the intestinal epithelium. *Gastroenterology* **139**, 904–917 (2010).
- Travassos, L. H. et al. Nod1 and Nod2 direct autophagy by recruiting ATG16L1 to the plasma membrane at the site of bacterial entry. *Nature Immunol.* **11**, 55–62 (2010).
- Cooney, R. et al. NOD2 stimulation induces autophagy in dendritic cells influencing bacterial handling and antigen presentation. *Nature Med.* **16**, 90–97 (2010).
- Saitoh, T. et al. Loss of the autophagy protein Atg16L1 enhances endotoxin-induced IL-1β production. *Nature* **456**, 264–268 (2008). **References 31 and 32 describe a link between key Crohn's disease susceptibility factors by showing that NOD2 can stimulate autophagy and that this constitutes an important bacterial handling mechanism.**
- Zhou, R., Yazdi, A. S., Menu, P. & Tschopp, J. A role for mitochondria in NLRP3 inflammasome activation. *Nature* **469**, 221–225 (2011).
- Martinon, F., Chen, X., Lee, A. H. & Glimcher, L. H. TLR activation of the transcription factor XBP1 regulates innate immune responses in macrophages. *Nature Immunol.* **11**, 411–418 (2010).
- Melmed, G. Y. & Targan, S. R. Future biologic targets for IBD: potentials and pitfalls. *Nature Rev. Gastroenterol. Hepatol.* **7**, 110–117 (2010).
- Maloy, K. J. & Kullberg, M. C. IL-23 and Th17 cytokines in intestinal homeostasis. *Mucosal Immunol.* **1**, 339–349 (2008).
- Kamada, N. et al. Unique CD14<sup>+</sup> intestinal macrophages contribute to the pathogenesis of Crohn disease via IL-23/IFN-γ axis. *J. Clin. Invest.* **118**, 2269–2280 (2008).
- Goodall, J. C. et al. Endoplasmic reticulum stress-induced transcription factor, CHOP, is crucial for dendritic cell IL-23 expression. *Proc. Natl Acad. Sci. USA* **107**, 17698–17703 (2010).
- Ahern, P. P. et al. Interleukin-23 drives intestinal inflammation through direct activity on T cells. *Immunity* **33**, 279–288 (2010).
- Leppkes, M. et al. RORγ-expressing Th17 cells induce murine chronic intestinal inflammation via redundant effects of IL-17A and IL-17F. *Gastroenterology* **136**, 257–267 (2009).
- Cosmi, L. et al. Human interleukin 17-producing cells originate from a CD161<sup>+</sup>CD4<sup>+</sup> T cell precursor. *J. Exp. Med.* **205**, 1903–1916 (2008).
- Kleinschek, M. A. et al. Circulating and gut-resident human Th17 cells express CD161 and promote intestinal inflammation. *J. Exp. Med.* **206**, 525–534 (2009).
- Yang, X. O. et al. Regulation of inflammatory responses by IL-17F. *J. Exp. Med.* **205**, 1063–1075 (2008).
- Chaudhry, A. et al. CD4<sup>+</sup> regulatory T cells control T<sub>H</sub>17 responses in a Stat3-dependent manner. *Science* **326**, 986–991 (2009).
- Buonocore, S. et al. Innate lymphoid cells drive interleukin-23-dependent innate intestinal pathology. *Nature* **464**, 1371–1375 (2010).
- Wu, S. et al. A human colonic commensal promotes colon tumorigenesis via activation of T helper type 17 T cell responses. *Nature Med.* **15**, 1016–1022 (2009).
- Ivanov, I. I. et al. Induction of intestinal Th17 cells by segmented filamentous bacteria. *Cell* **139**, 485–498 (2009).
- Gaboriau-Routhiau, V. et al. The key role of segmented filamentous bacteria in the coordinated maturation of gut helper T cell responses. *Immunity* **31**, 677–689 (2009). **References 47 and 48 document a strong link between segmented filamentous bacteria colonization and T<sub>H</sub>17 cells and, together with**



**reference 80, provide compelling evidence that colonization with distinct types of commensal bacterium results in the accumulation of different effector T cells in the intestine.**

49. Wolk, K., Witte, E., Witte, K., Warszawski, K. & Sabat, R. Biology of interleukin-22. *Semin. Immunopathol.* **32**, 17–31 (2010).
50. Pickert, G. *et al.* STAT3 links IL-22 signaling in intestinal epithelial cells to mucosal wound healing. *J. Exp. Med.* **206**, 1465–1472 (2009).
51. Sonnenberg, G. F. *et al.* Pathological versus protective functions of IL-22 in airway inflammation are regulated by IL-17A. *J. Exp. Med.* **207**, 1293–1305 (2010).
52. Chung, Y. *et al.* Critical regulation of early Th17 cell differentiation by interleukin-1 signaling. *Immunity* **30**, 576–587 (2009).
53. Ng, J. *et al.* *Clostridium difficile* toxin-induced inflammation and intestinal injury are mediated by the inflammasome. *Gastroenterology* **139**, 542–552.e3 (2010).
54. Muller, A. J. *et al.* The *S. Typhimurium* effector SopE induces caspase-1 activation in stromal cells to initiate gut inflammation. *Cell Host Microbe* **6**, 125–136 (2009).
55. Colonna, M. Interleukin-22-producing natural killer cells and lymphoid tissue inducer-like cells in mucosal immunity. *Immunity* **31**, 15–23 (2009).
56. Cua, D. J. & Tato, C. M. Innate IL-17-producing cells: the sentinels of the immune system. *Nature Rev. Immunol.* **10**, 479–489 (2010).
57. Park, S. G. *et al.* T regulatory cells maintain intestinal homeostasis by suppressing  $\gamma\delta$  T cells. *Immunity* **33**, 791–803 (2010).
58. Martin, B., Hirota, K., Cua, D. J., Stockinger, B. & Veldhoen, M. Interleukin-17-producing  $\gamma\delta$  T cells selectively expand in response to pathogen products and environmental signals. *Immunity* **31**, 321–330 (2009).
59. Sutton, C. E. *et al.* Interleukin-1 and IL-23 induce innate IL-17 production from  $\gamma\delta$  T cells, amplifying Th17 responses and autoimmunity. *Immunity* **31**, 331–341 (2009).
60. Spits, H. & Di Santo, J. P. The expanding family of innate lymphoid cells: regulators and effectors of immunity and tissue remodeling. *Nature Immunol.* **12**, 21–27 (2011).
61. Sawa, S. *et al.* Lineage relationship analysis of ROR $\gamma$ <sup>+</sup> innate lymphoid cells. *Science* **330**, 665–669 (2010).
62. Sonnenberg, G. F., Monticelli, L. A., Elloso, M. M., Fouser, L. A. & Artis, D. CD4<sup>+</sup> lymphoid tissue-inducer cells promote innate immunity in the gut. *Immunity* **34**, 122–134 (2011).
- References 45 and 62 identify new populations of LTI-like ILCs that secrete T<sub>H</sub>1 and T<sub>H</sub>17 pro-inflammatory cytokines in response to IL-23, and these contribute to intestinal pathology and host defences against intestinal pathogenic bacteria.**
63. Moro, K. *et al.* Innate production of T<sub>H</sub>2 cytokines by adipose tissue-associated c-Kit<sup>+</sup>Sca-1<sup>+</sup> lymphoid cells. *Nature* **463**, 540–544 (2010).
64. Saenz, S. A. *et al.* IL25 elicits a multipotent progenitor cell population that promotes T<sub>H</sub>2 cytokine responses. *Nature* **464**, 1362–1366 (2010).
65. Neill, D. R. *et al.* Nuocytes represent a new innate effector leukocyte that mediates type-2 immunity. *Nature* **464**, 1367–1370 (2010).
- References 63–65 describe various ILC populations that secrete T<sub>H</sub>2 cytokines (ILC type 2) in response to IL-25 and IL-33 and that can contribute to defence against intestinal helminth infection.**
66. Coombes, J. L. & Powrie, F. Dendritic cells in intestinal immune regulation. *Nature Rev. Immunol.* **8**, 435–446 (2008).
67. Varol, C., Zigmund, E. & Jung, S. Securing the immune tightrope: mononuclear phagocytes in the intestinal lamina propria. *Nature Rev. Immunol.* **10**, 415–426 (2010).
68. Varol, C. *et al.* Intestinal lamina propria dendritic cell subsets have different origin and functions. *Immunity* **31**, 502–512 (2009).
69. Bogunovic, M. *et al.* Origin of the lamina propria dendritic cell network. *Immunity* **31**, 513–525 (2009).
70. Schulz, O. *et al.* Intestinal CD103<sup>+</sup>, but not CX<sub>3</sub>CR1<sup>+</sup>, antigen sampling cells migrate in lymph and serve classical dendritic cell functions. *J. Exp. Med.* **206**, 3101–3114 (2009).
71. Niess, J. H. & Adler, G. Enteric flora expands gut lamina propria CX<sub>3</sub>CR1<sup>+</sup> dendritic cells supporting inflammatory immune responses under normal and inflammatory conditions. *J. Immunol.* **184**, 2026–2037 (2010).
72. Manicassamy, S. *et al.* Activation of  $\beta$ -catenin in dendritic cells regulates immunity versus tolerance in the intestine. *Science* **329**, 849–853 (2010).
73. Laffont, S., Siddiqui, K. R. & Powrie, F. Intestinal inflammation abrogates the tolerogenic properties of MLN CD103<sup>+</sup> dendritic cells. *Eur. J. Immunol.* **40**, 1877–1883 (2010).
74. Smith, P. D. *et al.* Intestinal macrophages and response to microbial encroachment. *Mucosal Immunol.* **4**, 31–42 (2011).
75. Murai, M. *et al.* Interleukin 10 acts on regulatory T cells to maintain expression of the transcription factor Foxp3 and suppressive function in mice with colitis. *Nature Immunol.* **10**, 1178–1184 (2009).
76. Hadis, U. *et al.* Intestinal tolerance requires gut homing and expansion of FoxP3<sup>+</sup> regulatory T cells in the lamina propria. *Immunity* **34**, 237–246 (2011).
- This study shows that oral tolerance requires homing and expansion of T<sub>reg</sub> cells in the intestine.**
77. Siddiqui, K. R., Laffont, S. & Powrie, F. E-cadherin marks a subset of inflammatory dendritic cells that promote T cell-mediated colitis. *Immunity* **32**, 557–567 (2010).
78. Platt, A. M., Bain, C. C., Bordon, Y., Sester, D. P. & Mowat, A. M. An independent subset of TLR expressing CCR2-dependent macrophages promotes colonic inflammation. *J. Immunol.* **184**, 6843–6854 (2010).
79. Izcue, A., Coombes, J. L. & Powrie, F. Regulatory lymphocytes and intestinal inflammation. *Annu. Rev. Immunol.* **27**, 313–338 (2009).
80. Atarashi, K. *et al.* Induction of colonic regulatory T cells by indigenous *Clostridium* species. *Science* **331**, 337–341 (2011).
- This paper describes a link between *Clostridium* spp. and T<sub>reg</sub> cells in the gut and, together with references 47 and 48, provides compelling evidence that colonization with distinct commensal bacteria results in the accumulation of different effector T cells in the intestine.**
81. Littman, D. R. & Rudensky, A. Y. Th17 and regulatory T cells in mediating and restraining inflammation. *Cell* **140**, 845–858 (2010).
82. Hall, J. A. *et al.* Commensal DNA limits regulatory T cell conversion and is a natural adjuvant of intestinal immune responses. *Immunity* **29**, 637–649 (2008).
83. Griseri, T., Asquith, M., Thompson, C. & Powrie, F. OX40 is required for regulatory T cell-mediated control of colitis. *J. Exp. Med.* **207**, 699–709 (2010).
84. Oldenhove, G. *et al.* Decrease of Foxp3<sup>+</sup> Treg cell number and acquisition of effector cell phenotype during lethal infection. *Immunity* **31**, 772–786 (2009).
85. Li, M. O. & Flavell, R. A. Contextual regulation of inflammation: a duet by transforming growth factor- $\beta$  and interleukin-10. *Immunity* **28**, 468–476 (2008).
86. Fantini, M. C. *et al.* Smad7 controls resistance of colitogenic T cells to regulatory T cell-mediated suppression. *Gastroenterology* **136**, 1308–1316.e3 (2009).
87. Pesu, M. *et al.* T-cell-expressed proprotein convertase furin is essential for maintenance of peripheral immune tolerance. *Nature* **455**, 246–250 (2008).
88. Perruche, S. *et al.* CD3-specific antibody-induced immune tolerance involves transforming growth factor- $\beta$  from phagocytes digesting apoptotic T cells. *Nature Med.* **14**, 528–535 (2008).
89. Torchinsky, M. B., Garaude, J., Martin, A. P. & Blander, J. M. Innate immune recognition of infected apoptotic cells directs T<sub>H</sub>17 cell differentiation. *Nature* **458**, 78–82 (2009).
90. Cong, Y., Feng, T., Fujihashi, K., Schoeb, T. R. & Elson, C. O. A dominant, coordinated T regulatory cell–IgA response to the intestinal microbiota. *Proc. Natl Acad. Sci. USA* **106**, 19256–19261 (2009).
91. Saraiva, M. & O'Garra, A. The regulation of IL-10 production by immune cells. *Nature Rev. Immunol.* **10**, 170–181 (2010).
92. Glocker, E. O. *et al.* Inflammatory bowel disease and mutations affecting the interleukin-10 receptor. *N. Engl. J. Med.* **361**, 2033–2045 (2009).
93. Smith, A. M. *et al.* Disordered macrophage cytokine secretion underlies impaired acute inflammation and bacterial clearance in Crohn's disease. *J. Exp. Med.* **206**, 1883–1897 (2009).
94. Winter, S. E. *et al.* Gut inflammation provides a respiratory electron acceptor for *Salmonella*. *Nature* **467**, 426–429 (2010).
95. Sartor, R. B. Microbial influences in inflammatory bowel diseases. *Gastroenterology* **134**, 577–594 (2008).
96. Frank, D. N. *et al.* Disease phenotype and genotype are associated with shifts in intestinal-associated microbiota in inflammatory bowel diseases. *Inflamm. Bowel Dis.* **17**, 179–184 (2011).
97. Willing, B. P. *et al.* A pyrosequencing study in twins shows that gastrointestinal microbial profiles vary with inflammatory bowel disease phenotypes. *Gastroenterology* **139**, 1844–1854.e1 (2010).
98. Maslowski, K. M. & Mackay, C. R. Diet, gut microbiota and immune responses. *Nature Immunol.* **12**, 5–9 (2011).
99. Cadwell, K. *et al.* Virus-plus-susceptibility gene interaction determines Crohn's disease gene *Atg16L1* phenotypes in intestine. *Cell* **141**, 1135–1145 (2010).
100. Garrett, W. S. *et al.* *Enterobacteriaceae* act in concert with the gut microbiota to induce spontaneous and maternally transmitted colitis. *Cell Host Microbe* **8**, 292–300 (2010).
- References 99 and 100 provide illustrative examples of how several host and environmental factors may act together to precipitate chronic intestinal inflammation.**

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank M. Asquith, H. Uhlig, P. Ahern and M. Barnes for review and G. Song-Zhao and O. Harrison for help with the figures. We apologize to those whose work was not cited owing to space constraints. K.J.M. and F.P. are supported by grants from the Wellcome Trust, Cancer Research UK and the European Union (FP7, INFLAMMACARE).

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at [www.nature.com/nature](http://www.nature.com/nature). Correspondence should be addressed to K.J.M. (kevin.maloy@path.ox.ac.uk) or F.P. (fiona.powrie@path.ox.ac.uk).

# Genetics and pathogenesis of inflammatory bowel disease

Bernard Khor<sup>1,2,3\*</sup>, Agnès Gardet<sup>1,2\*</sup> & Ramnik J. Xavier<sup>1,2,4</sup>

**Recent advances have provided substantial insight into the maintenance of mucosal immunity and the pathogenesis of inflammatory bowel disease. Cellular programs responsible for intestinal homeostasis use diverse intracellular and intercellular networks to promote immune tolerance, inflammation or epithelial restitution. Complex interfaces integrate local host and microbial signals to activate appropriate effector programs selectively and even drive plasticity between these programs. In addition, genetic studies and mouse models have emphasized the role of genetic predispositions and how they affect interactions with microbial and environmental factors, leading to pro-colitogenic perturbations of the host–commensal relationship.**

Inflammatory bowel disease (IBD) comprises the chronic relapsing inflammatory disorders Crohn's disease and ulcerative colitis. Family history is a risk factor for developing IBD, with a peak incidence in early adult life, although individuals of any age can be affected. IBD is thought to result from an inappropriate and continuing inflammatory response to commensal microbes in a genetically susceptible host. Recent progress in understanding IBD pathobiology offers insight into relevant disease mechanisms in mucosal immunity, including how genetic factors interact with microbial and environmental cues within tissue-specific contexts, the biological checkpoints involved, the selective decisions made during the course of disease and how plasticity of the biological response results in the capacity for different phenotypes.

Ulcerative colitis is characterized by inflammation that is limited to the colon: it begins in the rectum, spreads proximally in a continuous fashion and frequently involves the periappendiceal region. By contrast, Crohn's disease involves any part of the gastrointestinal tract — most commonly the terminal ileum or the perianal region — in a non-continuous fashion and, unlike ulcerative colitis, is commonly associated with complications such as strictures, abscesses and fistulas. Histologically, ulcerative colitis shows superficial inflammatory changes limited to the mucosa and submucosa with cryptitis and crypt abscesses. The microscopic features of Crohn's disease include thickened submucosa, transmural inflammation, fissuring ulceration and non-caseating granulomas.

Among complex diseases, genome-wide association studies (GWAS) have been successful in IBD, identifying 99 non-overlapping genetic risk loci, including 28 that are shared between Crohn's disease and ulcerative colitis<sup>1,2</sup> (Fig. 1). The genes implicated in childhood-onset and adult-onset IBD overlap, suggesting similar contributory genetic predispositions and pathophysiological pathways. Adding to the complexity of understanding disease mechanisms, a susceptibility allele often requires other genetic and non-genetic cues to manifest disease. The concordance rate in monozygotic twins of 10–15% in ulcerative colitis compared with 30–35% in Crohn's disease suggests that non-genetic factors may have an even more important role in ulcerative colitis than in Crohn's disease<sup>3</sup>. Furthermore, the higher penetrance of common Crohn's-disease-associated polymorphisms in genetic case-control studies than in population-based studies of

cohorts of the same ethnicity is probably due to the concomitant aggregation of both genetic and environmental factors in the case-control studies<sup>4</sup>. Smoking is an example of a disease-specific modifier that seems to exacerbate Crohn's disease while being protective against ulcerative colitis. Evidence suggests that smoking impairs autophagy, a process thought to be involved especially in Crohn's disease, demonstrating how exposure to a disease modifier in a genetically predisposed individual may mechanistically affect IBD development<sup>5</sup>.

In this Review, we provide an overview of genes and susceptibility loci implicated in IBD by GWAS and other genetic studies. Candidate genes are discussed in the context of IBD-relevant pathways, as well as how these molecular pathways interact with environmental factors to modulate intestinal homeostasis.

## Genes and pathways in IBD

International collaborative research groups focusing on an unbiased appraisal of the human genome have been particularly successful in identifying genes and genetic loci that contribute to IBD susceptibility<sup>1,2</sup>. Despite distinct clinical features, approximately 30% of IBD-related genetic loci are shared between ulcerative colitis and Crohn's disease, indicating that these diseases engage common pathways and may be part of a mechanistic continuum (Fig. 1).

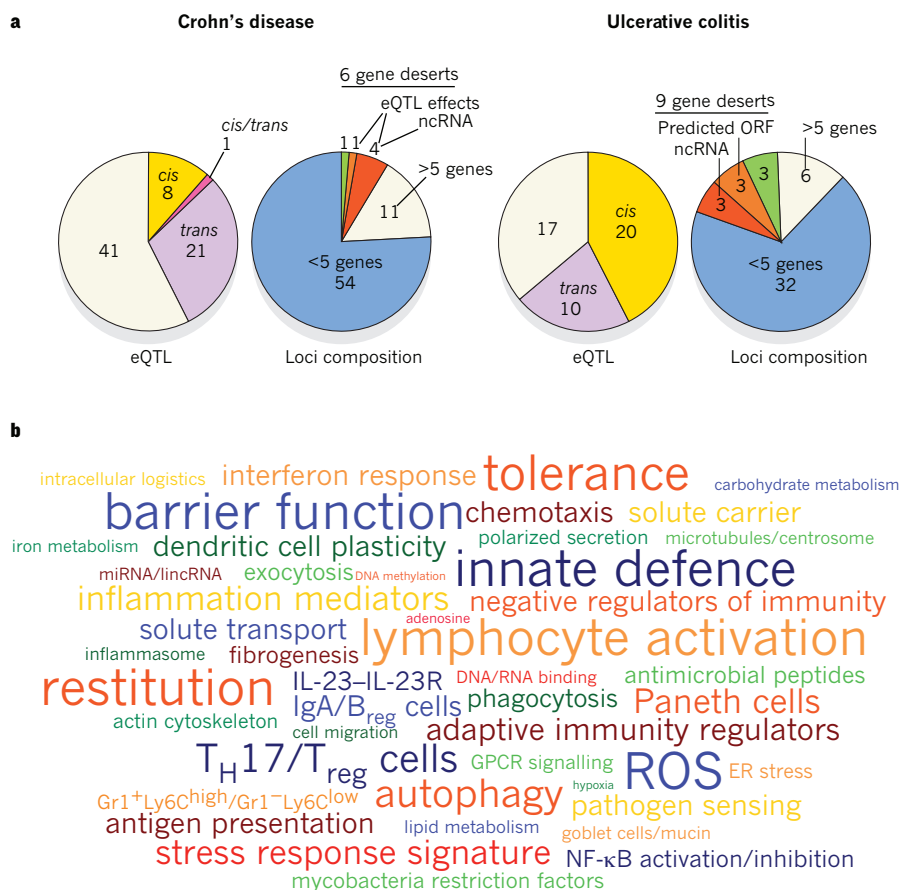
Analyses of the genes and genetic loci implicated in IBD show several pathways that are crucial for intestinal homeostasis, including barrier function, epithelial restitution, microbial defence, innate immune regulation, reactive oxygen species (ROS) generation, autophagy, regulation of adaptive immunity, endoplasmic reticulum (ER) stress and metabolic pathways associated with cellular homeostasis (Fig. 2). Early studies have suggested the existence of both protective and predisposing alleles<sup>6</sup>. Disease-relevant biological pathways are further highlighted when several components are implicated as risk factors together (Fig. 3).

Multidisease comparative analysis can uncover common disease-causing genes and pathways. More than 50% of IBD susceptibility loci have also been associated with other inflammatory and autoimmune diseases. These overlapping genes can have contrasting effects in different diseases. For example, the same coding variant of *PTPN22* (R620W) is a strong risk factor for type 1 diabetes and rheumatoid

<sup>1</sup>Gastrointestinal Unit and Center for the Study of Inflammatory Bowel Disease, Massachusetts General Hospital, Harvard Medical School, Boston, Massachusetts 02114, USA. <sup>2</sup>Center for Computational and Integrative Biology, Massachusetts General Hospital, Harvard Medical School, Boston, Massachusetts 02114, USA. <sup>3</sup>Pathology Service, Massachusetts General Hospital, Boston, Massachusetts 02114, USA. <sup>4</sup>The Broad Institute of MIT and Harvard, 7 Cambridge Center, Cambridge, Massachusetts 02142, USA.

\*These authors contributed equally to this work.





**Figure 1 | Genetic architecture of IBD-linked susceptibility loci.** **a**, GWAS have identified 71 risk loci in Crohn's disease and 47 risk loci in ulcerative colitis ( $P$  value of association  $< 5 \times 10^{-8}$ ). Of these, 28 risk loci exhibit shared associations (defined as  $P < 5 \times 10^{-8}$  for either Crohn's disease or ulcerative colitis, and  $P < 1 \times 10^{-4}$  for the other form of IBD). Approximately half of the loci implicated in Crohn's disease and ulcerative colitis are associated with *cis*- and/or *trans*-expression quantitative trait loci (eQTL) effects (left panels). Genes whose expression are affected by these variants could also be involved in IBD pathogenesis. The loci composition (right panels) shows the number of genes that either lie within or segregate in linkage disequilibrium with IBD-implicated loci (coefficient of correlation  $r^2 > 0.8$ ). These loci are structurally heterogeneous, and are associated with widely ranging numbers of genes. Loci not associated with any genes, known as gene deserts, frequently contain non-coding transcripts or predicted open reading frames (ORFs), and can be associated with *trans*-eQTL effects. **b**, Recurring terms illustrating biological processes implicated by at least three genes represented in IBD loci; font sizes are proportional to the number of genes associated with each respective process. B<sub>reg</sub> cells, B regulatory cells; ER, endoplasmic reticulum; GPCR, G-protein-coupled receptor; IL, interleukin; lincRNA, large intervening non-coding RNA; miRNA, microRNA; ncRNA, non-coding RNA; NF-κB, nuclear factor-κB; ROS, reactive oxygen species; T<sub>H</sub>17 cells, T helper 17 cells; T<sub>reg</sub> cells, T regulatory cells.

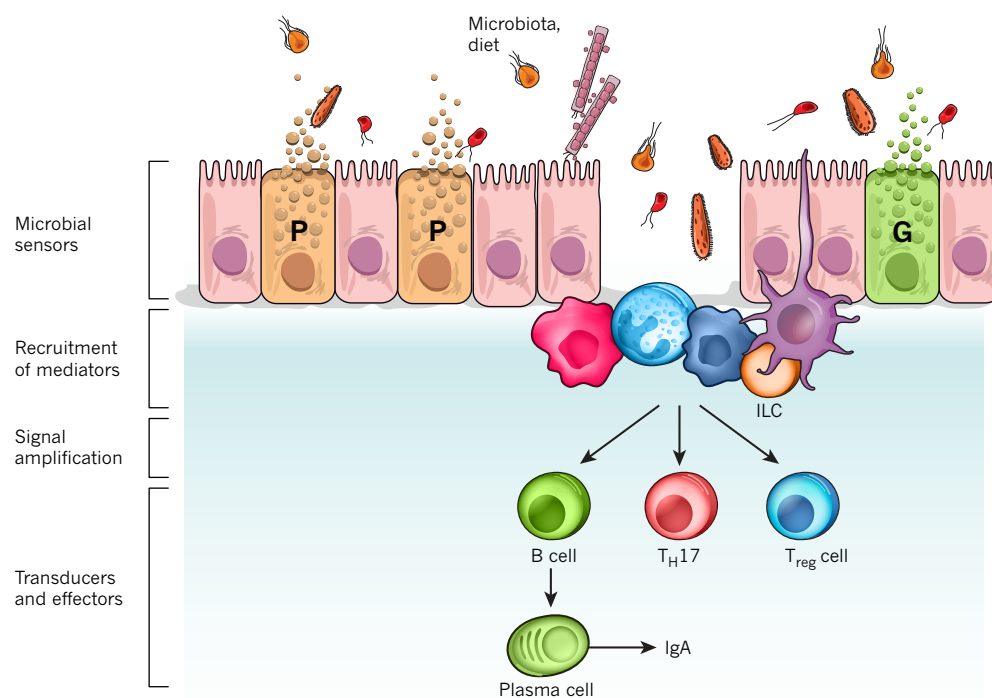
arthritis, but is protective against Crohn's disease<sup>7</sup>. These data suggest that crucial clues to disease biology may reside in understanding the function of these shared genes. Several loci containing genes such as *MST1*, *IL2*, *CARD9* and *REL* are shared between ulcerative colitis and the associated complication primary sclerosing cholangitis (PSC)<sup>8</sup>. This overlap may help to identify subsets of patients with ulcerative colitis who are at risk of PSC. Risk loci for Crohn's disease present an unexpected overlap with susceptibility regions for *Mycobacterium leprae* infection, including genes such as *NOD2*, *C13orf31* and *LRRK2* (ref. 9). Although absent from the leprosy GWAS, other Crohn's-disease-associated genes are also implicated in host responses to mycobacterial infection, including *CARD9*, *LTA*, *ITLN1* and *IRGM*<sup>1</sup>. Thus, studies to delineate immune responses to antigens from, and infection by, mycobacteria, or other microbes that elicit similar host cell responses, may also be pertinent to Crohn's disease.

Genetic variants associated with IBD can vary in frequency depending on the cohort ethnicity, raising the possibility that some such variants may have emerged in the context of historical selective pressures. Although this notion remains to be demonstrated in IBD, lessons from other autoimmune and infectious contexts lend support. For example, variants of apolipoprotein L1 and the inhibitory Fc receptor FcγRIIb that confer protection against trypanosomiasis and malaria, respectively, are more common in populations endemically exposed to these pathogens, but these variants also confer increased susceptibility to focal segmental glomerulosclerosis and systemic lupus erythematosus (SLE), respectively<sup>10,11</sup>.

Current GWAS are typically powered to characterize variants of  $>1\%$  frequency and do not include the contributions from rare variants ( $<1\%$  frequency). Exome sequencing can be useful for identifying rare variants, whereas whole-genome sequencing is of value in elucidating modifier loci. If pedigrees are available, rare variant discovery can be further targeted by fine mapping, as shown by the identification of *IL10RA* polymorphisms associated with the development of

early-onset IBD<sup>12</sup>. Other interleukin-10 receptor (IL-10R) signalling components have also been implicated by GWAS, including *STAT3*, *TYK2*, *JAK2* and *IL10* itself, in concordance with the notion that both rare and common variants may highlight the same pathway. Although these components can also function in other contexts — for example, the transcription factor *STAT3* and the kinase proteins *TYK2* and *JAK2* are involved in the signalling of the interleukins IL-6, IL-22 and IL-23 — these results illustrate the value of genetic studies in determining not just single genes, but also disease-relevant pathways. Recent resequencing studies in IBD recovered both known and new variants of *CARD9*, *NOD2* and *IL23R*, with independent effects on disease risk. The *IL23R* variants were protective, supporting previous findings of a common protective *IL23R* allele and illustrating how studies of rare variants can reinforce GWAS findings<sup>6</sup>. Furthermore, T helper 17 (T<sub>H</sub>17) cells generated *ex vivo* from subjects with a variant *IL23R* allele (R381Q) show decreased production of the pro-inflammatory cytokine IL-17A in response to IL-23 stimulation, emphasizing the importance of IL-23-related pathways in human IBD<sup>13</sup>.

Early functional studies attempting to determine causality have largely focused on coding variants, although non-coding single nucleotide polymorphisms (SNPs) can be associated with qualitative and quantitative changes. Alternative splicing exemplifies a qualitative change affected by non-coding modifications. In the context of regulating immune responses, *IL23R* and *NOD2* can encode truncated variants that inhibit their signalling pathways<sup>14,15</sup>. Furthermore, genetic changes may affect transcription-factor-binding sequences, locus accessibility, translational efficiency and *trans*-regulators such as non-coding RNAs and microRNAs (miRNAs). In this regard, a Crohn's-disease-associated synonymous variant in *IRGM* (c.313C>T) perturbs regulation by miR-196A and miR-196B, and is associated with altered *IRGM* expression in patients with Crohn's disease who bear this SNP<sup>16</sup>. *Cis*- or *trans*-expression quantitative trait loci (eQTL) are detected for approximately half of the IBD risk regions,



### Cellular responses

#### Autophagy

ATG16L1\*, IRGM, NOD2\*, LRRK2, CUL2, PARK7, DAP

#### Apoptosis/necroptosis

FASLG, THADA\*, DAP, PUS10, MST1\*

#### ER stress

CPEB4, ORMDL3, SERINC3, XBP1\*

#### Carbohydrate metabolism

GCKR\*, SLC2A4RG

#### Intracellular logistics

VAMP3, KIF21B, TTL8, FGFR10P, CEP72, TPPP

#### Oxidative stress

PRDX5, BACH2, ADO, GPX4, GPX1\*, SLC22A4, LRRK2, NOD2\*, CARD9\*, HSPA6, DLD, PARK7, UTS2\*, PEX13

#### Cell migration

ARPC2, LSP1, AAMP

### IBD-related processes

#### Epithelial barrier

GNA12\*, HNF4A, CDH1, ERRF1, MUC19, ITLN1\*

#### Restitution

REL, PTGER4, NKX2-3, STAT3, ERRF1, HNF4A, PLA2G2A/E

#### Solute transport

SLC9A4, SLC22A5, SLC22A4\*, AQP12A/B, SLC9A3, SLC26A3

#### Paneth cells

ITLN1\*, NOD2\*, ATG16L1\*, XBP1\*

#### Innate mucosal defence

NOD2\*, ITLN1\*, CARD9\*, REL, SLC11A1, FCGR2A\*/B

#### Immune cell recruitment

CCL11/CCL2/CCL7/CCL8, CCR6, IL8RA/IL8RB, MST1\*

#### Antigen presentation

ERAP2\*, LNPEP, DENND1B

#### IL-23/T<sub>H</sub>17

IL23R\*, JAK2, TYK2\*, STAT3, ICOSLG, IL21, TNFSF15\*

#### T-cell regulation

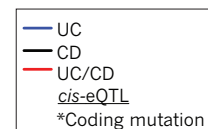
NDFIP1, TNFSF8, TAGAP, IL2, IL2R $\alpha$ , TNFRSF9, PIM3, IL7R\*, IL12B, IL23, PRDM1, ICOSLG, TNFSF8, IFNG, IL2

#### B-cell regulation

IL5, IKZF1, BACH2, IL7R\*, IRF5

#### Immune tolerance

IL10, IL27\*, SBNQ2, CREM, IL1R1/IL1R2, NOD2\*



**Figure 2 | A model for IBD pathways based on GWAS.** Intestinal homeostasis involves the coordinated actions of epithelial, innate and adaptive immune cells. Barrier permeability permits microbial incursion, which is detected by the innate immune system, which then orchestrates appropriate tolerogenic, inflammatory and restitutive responses in part by releasing extracellular mediators that recruit other cellular components, including adaptive immune cells. Genetic variants, the microbiota and immune factors

affect the balance of these signals. Genes in linkage disequilibrium ( $r^2 > 0.8$ ) with IBD-associated single nucleotide polymorphisms (SNPs) were manually curated and classified according to their function(s) in the context of intestinal homeostasis and immunity. Text colour indicates whether the genes are linked to risk loci associated with Crohn's disease (CD; black), ulcerative colitis (UC; blue) or both (red). Asterisk denotes corresponding coding mutations; *cis*-eQTL effects are underlined. G, goblet cell; P, Paneth cell.

indicating that allele-specific gene-expression changes contribute to disease risk (Fig. 1). Furthermore, IBD-implicated loci contain more than 10 miRNA-encoding sequences and 39 large intervening non-coding RNAs (lincRNAs), 5 of which interacted with the histone methyltransferase polycomb repressive complex 2 (PRC2), supporting the notion that regulation of gene expression by miRNAs and lincRNAs may be mechanistically relevant in IBD<sup>17</sup>.

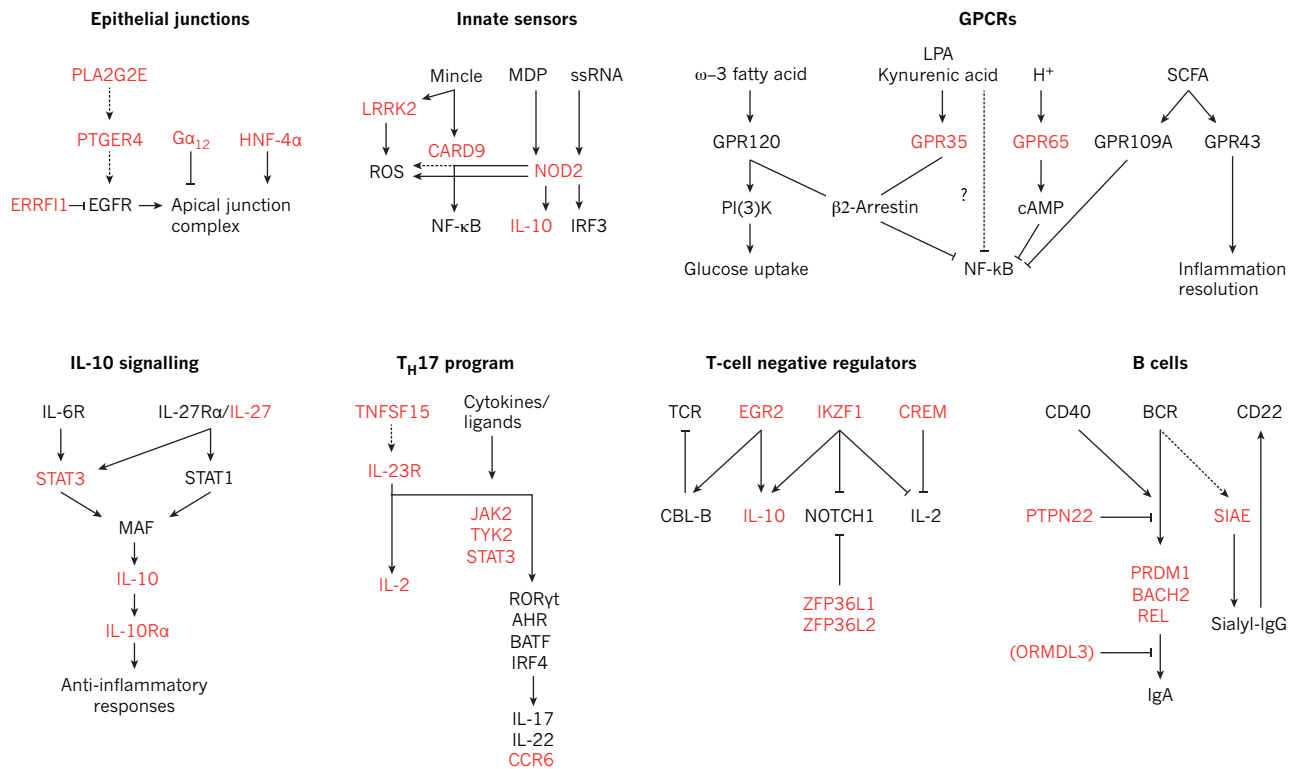
So far, GWAS account for 23% and 16% of the heritability in Crohn's disease and ulcerative colitis, respectively<sup>1,2</sup>. Although these may be underestimates owing to the net effect of common variants that are individually too small to calculate accurately; the missing heritability may further comprise genetic, epigenetic and non-genetic (including environmental) components. Genetic factors such as rare variants, private mutations, structural variants and interactions between genes are not well captured by GWAS. Nevertheless, a key success of GWAS in IBD has been the ability to provide insight into disease pathobiology by highlighting key molecular pathways.

### Epithelial encounters and pathogenicity

The intestinal mucosa exists in a functional equilibrium with the complex luminal milieu, which is dominated by a spectrum of microbial species and their products. Maintaining this functional balance is central to preserving normal mucosal physiology, with perturbations contributing to the pathophysiology of many gastrointestinal disorders, including IBD. In addition to nutrient absorption, intestinal epithelial cells (IECs) perform both barrier and signal-transduction functions, with the capacity to sense luminal contents through surface receptors and, in return, secrete regulatory products that can orchestrate an appropriate response in the underlying lamina propria.

Molecular details of the epithelial barrier and the structure of tight junctions, which are crucial to its integrity, have been characterized. Abnormal intestinal permeability has been observed in IBD patients and in some of their first-degree relatives. Genes within several IBD-associated loci indicate a role for barrier integrity in disease predisposition, implicating candidate genes such as *CDH1*, *GNA12* and





**Figure 3 | Genetic variants in IBD signalling modules.** Schematic of selected signalling pathways involved in the maintenance of intestinal homeostasis, including epithelial junctional complex assembly, innate immune recognition of pathogen-associated motifs, GPCRs and immune defence, anti-inflammatory interleukin-10 (IL-10) signalling,  $T_H17$ -cell differentiation, inhibitory pathways in lymphocyte signalling, and B-cell activation and IgA antibody responses. Proteins encoded by genes identified as being in linkage disequilibrium with IBD-risk SNPs ( $r^2 > 0.8$ ) are highlighted in red.

*PTPN2*. Genetic studies have shown that truncated forms of the adherens junction protein E-cadherin (encoded by *CDH1*) are associated with Crohn's disease, and intestinal biopsies from patients with Crohn's disease carrying these mutant alleles show inappropriate protein localization and cytosolic accumulation<sup>18</sup>. Activation of the G protein  $G_{\alpha_{12}}$  (encoded by *GNA12*) leads to phosphorylation of the tight junction proteins ZO-1 and ZO-2, resulting in destabilization of cell junctions in epithelial cell lines<sup>19</sup>. *In vitro* studies show that the protein tyrosine phosphatase family member *PTPN2* protects against interferon- $\gamma$  (IFN- $\gamma$ )-induced epithelial permeability; concordantly, *Ptpn2*-deficient mice show increased susceptibility to experimental colitis<sup>20,21</sup>.

Genetic studies have associated IBD with several transcription factors involved in epithelial regeneration, such as *HNF4A* and *NKX2-3*, which control crypt cell proliferation and IEC differentiation, respectively<sup>22–24</sup>. Spontaneous colitis did not occur in all animal models with IEC-specific deletion of *Hnf4a*, suggesting that further environmental triggers are required for disease<sup>22,23</sup>. *STAT3*, the gene encoding which lies within an IBD-implicated locus, is activated in epithelial cells from patients with IBD, and IEC-specific *Stat3* deletion affects epithelial repair<sup>25</sup>.

The intestinal barrier is enhanced by the presence of a pre-epithelial layer formed primarily of mucus glycoproteins, trefoil peptides, IgA and antimicrobial peptides (AMPs). Goblet cells generate the mucus layer, a protective polysaccharide bilayer rich in cationic proteins, the inner layer of which is essentially devoid of microbes. Patients with IBD frequently have a compromised mucus layer and increased mucolytic bacteria; mucus layer defects are also observed in *Muc2*<sup>-/-</sup> and IEC-specific *C1galt1*<sup>-/-</sup> mice, which develop spontaneous colitis<sup>26</sup>. Interestingly, some patients with ulcerative colitis show defective intestinal O-glycosylation resembling that seen in *C1galt1*<sup>-/-</sup> mice<sup>26</sup>. Paneth cells are located in the crypts of the small intestine. In addition

to the role of these cells in crypt homeostasis and maintenance of the intestinal stem-cell niche, they also secrete antimicrobial effectors that prevent microbial invasion and control the composition of the gut microflora. These effectors include lysozyme, RegIII $\gamma$ , secreted phospholipase A<sub>2</sub> (which degrades bacterial membrane phospholipids) and defensins HD5 and HD6 (pore-forming hydrophobic peptides that can integrate into bacterial membranes, resulting in lysis). Production of AMPs is regulated by Toll-like receptor (TLR) and NOD2 signals triggered by commensal flora. Paneth cell defects and susceptibility to intestinal inflammation have been uncovered in mice deficient in several Crohn's-disease-associated genes, including *Nod2*, *Atg16l1* and *Xbp1* (refs 27–29). These results highlight pathways important to Paneth cell biology, such as the regulation of AMP production (*Nod2*), granule exocytosis (*Atg16l1*) and the ER stress response (*Xbp1*). Similar phenotypes have been observed in human disease, such that patients with Crohn's disease carrying the *ATG16L1* (T300A) mutation show Paneth cell granule abnormalities. These findings suggest that defects in Paneth cell biology may define a subset of patients with Crohn's disease.

Cells with high synthetic capacity and secretory activity, such as Paneth cells and goblet cells, have high baseline levels of ER stress, leading to activation of the unfolded protein response (UPR), which controls cellular programs that allow proper protein processing. The UPR is mainly cytoprotective, although it can signal apoptosis after sustained ER stress. Increased intestine epithelial ER stress and susceptibility to colitis have been observed in mice with overactivation of, or perturbations in, the UPR pathway, including *Muc2* missense mutation, *Agr2*<sup>-/-</sup>, *Ern2*<sup>-/-</sup> (also known as *Ire1b*<sup>-/-</sup>), IEC-specific *Xbp1*<sup>-/-</sup> and *Mbtps1*-hypomorphic mice<sup>29,30</sup>. Similarly, studies in primary IECs from patients with IBD show activated ER stress responses,

leading to activation of the unfolded protein response (UPR), which controls cellular programs that allow proper protein processing. The UPR is mainly cytoprotective, although it can signal apoptosis after sustained ER stress. Increased intestine epithelial ER stress and susceptibility to colitis have been observed in mice with overactivation of, or perturbations in, the UPR pathway, including *Muc2* missense mutation, *Agr2*<sup>-/-</sup>, *Ern2*<sup>-/-</sup> (also known as *Ire1b*<sup>-/-</sup>), IEC-specific *Xbp1*<sup>-/-</sup> and *Mbtps1*-hypomorphic mice<sup>29,30</sup>. Similarly, studies in primary IECs from patients with IBD show activated ER stress responses,

and hypomorphic variants of XBP1 have been associated with risk of IBD<sup>29</sup>. Overall, these results indicate that genetic variants that perturb mechanisms that protect against ER stress can affect intestinal homeostasis in IBD. In addition to its effects on cell viability, ER stress also activates autophagy and IL-23 release, suggesting that sustained ER stress may engage inflammatory circuits that are subsequently propagated by T cells<sup>31</sup>.

In addition to limiting bacterial translocation across the mucosal barrier, IECs promote intestinal homeostasis by regulating innate and adaptive immune responses. Illustrating this point, IECs produce intestinal alkaline phosphatase, which can mediate lipopolysaccharide detoxification. Resolvin-E1, which is generated in part through the action of epithelial cyclooxygenase-2, attenuates neutrophil transmigration and upregulates epithelial expression of intestinal alkaline phosphatase during the restitutive response, a process termed epithelial imprinting<sup>32</sup>. IECs can also modulate adaptive immune responses, driving the differentiation of anti-inflammatory T regulatory (T<sub>reg</sub>) cells by releasing the vitamin A metabolite retinoic acid and the cytokines thymic stromal lymphopoietin (TSLP) and transforming growth factor- $\beta$  (TGF- $\beta$ )<sup>33</sup>. Breakdown in such epithelial defence mechanisms could lead to pathological intestinal inflammation.

### Checkpoints in the innate immune response

The physical barrier of the intestinal epithelium is complemented by a well-evolved mucosal innate immune system, which is populated by cells poised to defend against pathogenic incursions and curtail inflammatory responses to maintain a state of hyporesponsiveness to commensal bacteria. Dendritic cells, macrophages, innate lymphoid cells (ILCs) and neutrophils are crucial cellular components of the innate immune system during infection or inflammation. Supporting the notion that defective innate immune responses can lead to IBD, patients with innate immunodeficiencies such as chronic granulomatous disease and Hermansky–Pudlak syndrome, which is associated with defective responses to bacterial DNA motifs (CpG oligonucleotides) specifically in plasmacytoid dendritic cells, tend to develop IBD<sup>34</sup>. Similarly, patients with Crohn's disease have defective innate immune responses, including attenuated macrophage activity *in vitro*, and impaired neutrophil recruitment and exogenous *Escherichia coli* clearance *in vivo*<sup>35</sup>.

Intestinal dendritic cells constitute a central interface for monitoring the environment and relaying signals to initiate appropriate adaptive immune responses<sup>33</sup>. Dendritic cell subsets are specialized and respond to endogenous and exogenous stimuli such as microbial motifs, fatty acids, oxidized lipids and vitamin D by selectively engaging pro-inflammatory, anti-inflammatory, epithelial restitutive or T-cell education programs, as well as inducing IgA production<sup>33,36</sup>. For example, T<sub>reg</sub>-cell differentiation can be promoted by tolerogenic dendritic cells induced by TSLP, TGF- $\beta$  and retinoic acid, all of which are made by IECs and stromal cells; these dendritic cells express the integrin CD103 but not the chemokine receptor CX<sub>3</sub>CR<sub>1</sub> (ref. 33). By contrast, dendritic cells expressing E-cadherin are a pro-inflammatory subset that promotes T<sub>H</sub>17-cell differentiation (see ref. 37 and page 298 for further details). Bacterial flagellins can override dendritic cell tolerogenic programs by stimulating TLR5 and inducing the release of pro-inflammatory mediators from hyporesponsive lamina propria CD11c<sup>high</sup> dendritic cells, pointing to a broader role for flagellated bacteria in IBD<sup>38</sup>. This specific immunostimulatory role for TLR5 may be particularly relevant in IBD, as seroreactivity to the bacterial flagellin CBir1, observed in approximately 50% of patients with Crohn's disease, correlates with a complicated clinical course.

Intestinal homeostasis is maintained in part by the actions of resident macrophages that have enhanced phagocytic and bactericidal activity and decreased production of pro-inflammatory cytokines. Specialized macrophage subsets are also involved; tumour-necrosis factor- $\alpha$  (TNF- $\alpha$ )-secreting and IL-1 $\beta$ -secreting Ly6C<sup>high</sup> monocytes are recruited in the initial phase of microbial challenge or tissue injury, whereas reparative IL-10-secreting, TGF- $\beta$ -secreting and

vascular-endothelial-growth-factor-secreting Ly6C<sup>low</sup> monocytes are mobilized during the resolution phase of inflammation<sup>39</sup>. Neutrophils may also contribute to the resolution of inflammation, for example, by synthesizing anti-inflammatory mediators such as lipoxin A<sub>4</sub>. Studies showing impaired secretion of lipoxin A<sub>4</sub> in mucosal tissues from patients with ulcerative colitis support the relevance of such mechanisms in IBD.

IL-22 is emerging as an important cytokine in epithelial homeostasis, showing protective activity in different models of colitis through its stimulatory effect on antimicrobial and reparative processes. Produced by several cell types, such as ILCs, lymphoid tissue induced (LTi) cells, T<sub>H</sub>17 cells and  $\gamma\delta$  T cells, most intestinal IL-22 at steady state is produced by ILCs expressing the transcription factor ROR $\gamma$ <sup>40,41</sup>. Studies in patients with Crohn's disease have shown decreased frequencies of IL-22-secreting ILCs in the lamina propria<sup>42</sup>. Together, these findings suggest a central role for ILCs (and other IL-22-producing cells) in regulating intestinal homeostasis, which remains to be characterized in IBD.

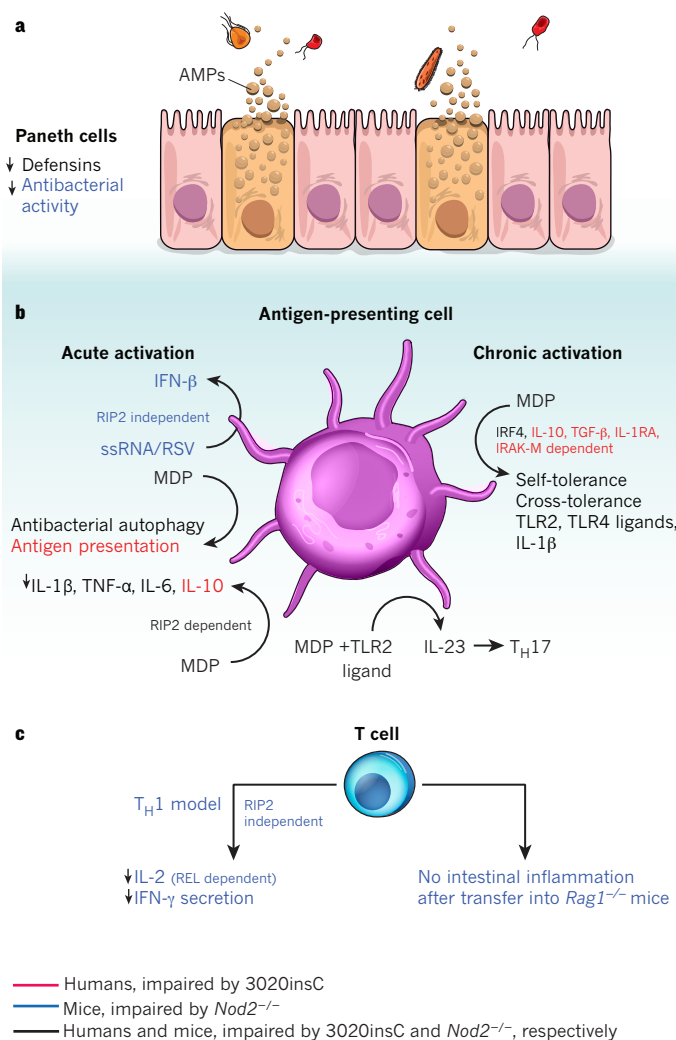
### NOD2 and IBD

NOD2 was the first gene to be associated with IBD, and thereafter several genes that interact epistatically with NOD2 signalling were also implicated. NOD2 recognizes the peptidoglycan product muramyl dipeptide (MDP), which modulates both innate and adaptive immune responses<sup>43</sup> (Fig. 4). For example, MDP stimulation induces autophagy, which controls bacterial replication and antigen presentation, and acts on dendritic cells in conjunction with TLR ligands to promote T<sub>H</sub>17-cell differentiation<sup>44,45</sup>. NOD2 may also contribute to immune tolerance. These effects are impaired in cells from patients with the Crohn's-disease-associated NOD2 mutation 3020insC. Furthermore, NOD2 can participate in distinct MDP-independent pathways such as regulation of the T-cell response and the type I IFN response to single-stranded RNA (ssRNA) stimulation, indicating that gut microbial ssRNAs may exist and have immunomodulatory properties<sup>46</sup>. The relative contributions of these cytosolic MDP-sensing pathways vary greatly between cell types (Fig. 4). Further studies are needed to uncover the effect of disease-associated NOD2 alleles in different cell-specific programs, and unravel the precise role(s) of NOD2 in IBD. Other families of innate immune receptors linked to intestinal inflammation and immunity include NOD-like receptors (NLRs) and RIG-I-like receptors (RLRs). These receptors recognize microbial motifs or damage-associated molecular patterns and can activate the inflammasome, thus appropriate regulation of these pathways is required for intestinal homeostasis. For example, mouse knockout studies of *Nlrp3* or *RIG-I* (also known as *Ddx58*) show increased susceptibility to experimental colitis<sup>47</sup>. Conversely, sustained overactivation of NLRs can also have detrimental effects, as illustrated by activating mutations in NOD2 and NLRP3 giving rise to Blau syndrome and cryopyrinopathies, respectively.

### CARD9 and IBD

CARD9 is an IBD-implicated adaptor protein that integrates signals from many innate immune receptors that recognize viral, bacterial and fungal motifs. Depending on the stimulus, CARD9 interacts with distinct signalling complexes and activates different pathways to modulate cytokine environments appropriately<sup>48,49</sup>. In particular, recognition of fungal motifs in human dendritic cells leading to CARD9 and dectin-1 signalling results in the broad activation of members of the nuclear factor- $\kappa$ B (NF- $\kappa$ B) transcription factor family, whereas CARD9 and dectin-2 signalling selectively activates the IBD-implicated NF- $\kappa$ B factor REL, enhancing the production of T<sub>H</sub>17-polarizing cytokines such as IL-1 $\beta$  and the IL-23 p19 subunit<sup>50</sup>. Defective CARD9 function leads to the immune disorder mucocutaneous candidiasis, at least in part owing to failure to promote an adequate T<sub>H</sub>17 immune response. These data illustrate how innate immune signalling molecules, including NOD2 and CARD9, can act as central hubs to integrate diverse signals and





**Figure 4 | Cell-intrinsic functions of NOD2.** NOD2 is activated by the bacterial peptidoglycan muramyl dipeptide (MDP). Cell-specific NOD2 functions are shown, distinguishing between those functions impaired in cells from humans with the Crohn's-disease-associated mutation 3020insC (red), from *Nod2*-deficient mice (blue), or from both (black). **a**, In Paneth cells, *Nod2* deficiency leads to attenuated antibacterial activity in the intestinal crypts and decreased expression of  $\alpha$ -defensin 4 (encoded by *Defcr4*, also known as *Defa4*) and  $\alpha$ -defensin-related sequence 10 (DEFSCR-RS10, also known as DEFA-RS10). **b**, MDP-stimulated release of pro-inflammatory NF- $\kappa$ B-dependent cytokines (such as IL-1 $\beta$ , TNF- $\alpha$  and IL-6), as well as secretion of IL-23 (which promotes T<sub>H</sub>17 differentiation) after co-stimulation with MDP and TLR2 ligands, is decreased in antigen-presenting cells from *Nod2*-deficient mice or 3020insC human donors. MDP stimulation also leads to NOD2-activated autophagy and antigen presentation. In mice, the activation of antigen-presenting cells by ssRNA or respiratory syncytial virus (RSV) stimulates secretion of type I interferon (IFN- $\beta$ ) in a NOD2-dependent, receptor-interacting protein-2 (RIP2)-independent fashion. In contrast to the pro-inflammatory effects, chronic NOD2 activation (right) by MDP induces both self-tolerance and cross-tolerance to IL-1 $\beta$ , and TLR2 and TLR4 ligands. This is dependent on IRF4 in mice and humans, and also on IL-10, TGF- $\beta$ , IL-1RA and IL-1R-associated kinase M (IRAK-M) in humans. MDP-induced tolerance is lost in *Nod2*-deficient mice and in patients with the 3020insC variant. NOD2-dependent release of IL-10 after MDP stimulation has been demonstrated to be specific to humans and is impaired in 3020insC cells. MDP-stimulated release of several cytokines, including IL-10, IL-1 $\beta$ , TNF- $\alpha$  and IL-6, is dependent on RIP2. **c**, In mice, NOD2 mediates IFN- $\gamma$  secretion and REL-dependent IL-2 production in T cells in response to *Toxoplasma gondii* infection. Also, *Nod2* deficiency attenuates the ability of T cells to cause experimental colitis after transfer into *Rag1*-deficient hosts.

selectively activate specific effector pathways; in the polymicrobial context of the gut, it seems reasonable that defects at such nodal points would constitute key predispositions to IBD.

### Redox equilibrium in IBD

The reduction and oxidation (redox) state of the gut depends on an equilibrium between oxidants, such as free radicals, ROS or reactive nitrogen species, and antioxidant mechanisms, such as the glutathione peroxidase (GPX) and glutathione S-transferase enzymes. This redox state affects many signal-transduction pathways, such as NF- $\kappa$ B signalling and AMP activity<sup>51</sup>. Supporting the importance of antioxidant pathways in intestinal homeostasis, mice deficient in both *Gpx1* and *Gpx2* develop spontaneous colitis. IBD genetic studies have implicated loci containing *GPX1* and *GPX4*, further highlighting the relevance of these mechanisms in disease (Fig. 2). Among the oxidants, ROS represent an important class of effector molecules generated by mitochondrial and non-mitochondrial sources. ROS are non-toxic at basal levels and are even required to maintain the intestinal stem-cell niche. In the context of innate immunity, ROS have important antimicrobial activity, and contribute to intracellular signalling, promoting the production of pro-inflammatory cytokines. Furthermore, ROS generated by epithelial cells after infection can transmit signals to adjacent cells in a paracrine manner, allowing the local coordination of chemokine production<sup>52</sup>. Genes within several IBD-associated loci may either regulate ROS production or protect against oxidative stress (Fig. 2). In particular, NOD2, CARD9 and IFN- $\gamma$ -regulated leucine-rich repeat kinase 2 (LRRK2) all contribute to ROS production<sup>43,53,54</sup>. In addition to pro-inflammatory pathways, ROS are also involved in T<sub>reg</sub>-cell polarization and function<sup>55,56</sup>. Thus, understanding the role of disease variants will require a broader understanding of the cell- and tissue-specific effects of ROS.

### Autophagy and IBD

Genetic analyses have shown an unsuspected role for autophagy in innate immunity and IBD, implicating two component genes, *ATG16L1* and *IRGM*, in IBD pathogenesis<sup>57–59</sup>. Autophagy is involved in intracellular homeostasis, contributing to the degradation and recycling of cytosolic contents and organelles, as well as to resistance against infection and the removal of intracellular microbes (Fig. 5). *ATG16L1* is essential for all forms of autophagy, and the coding mutation T300A is associated with increased risk of Crohn's disease. Despite ubiquitous expression of *ATG16L1*, defects associated with *ATG16L1* polymorphisms have so far been described only within the gut, probably owing to the high microbial load in this tissue. Subsequent evidence for MDP stimulation of NOD2-activated autophagy illustrates a link between genetic risk loci, and highlights the importance of defining disease-associated pathways and the potential of new roles for known genes<sup>44,45</sup>. Epithelial cells and dendritic cells containing Crohn's-disease-associated *ATG16L1* and *NOD2* variants show defects in antibacterial autophagy<sup>44,45,60</sup>. In dendritic cells, these defects are associated with an impaired ability to present exogenous antigens to CD4<sup>+</sup> T cells<sup>44</sup>. These results illustrate a close relationship between NOD2, *ATG16L1* and autophagy, affecting intracellular processing and communication with the adaptive immune system, suggesting that genetic polymorphisms may affect both pathways concomitantly.

Abnormalities consistent with Crohn's disease have been observed in mice with defects in autophagy, including hypomorphic *Atg16l1* (*Atg16l1*<sup>HM</sup>) and IEC-specific *Atg5*-deficient mice<sup>27</sup>. Paneth cells either from *Atg16l1*<sup>HM</sup> mice or from patients with Crohn's disease who have the *ATG16L1* (T300A variant) allele show aberrant granule size, number and location, and reduced AMP secretion; notably, they also show gain of function, as evidenced by upregulated peroxisome proliferator-activated receptor signalling<sup>27</sup>. The landmark findings that gnotobiotic (germ-free) *Atg16l1*<sup>HM</sup> mice lost these Paneth cell anomalies and their sensitivity to dextrate sulphate sodium (DSS)-induced colitis, and that these abnormalities were restored by norovirus infection provide a definitive demonstration of how host-microbial interactions contribute to the pathophysiology of IBD<sup>61</sup>.

## Effectors and regulators of adaptive immunity

Homeostasis in the gut involves a balance between anti-inflammatory and pro-inflammatory signals, such that inflammatory disease results from an inadequate  $T_{reg}$ -cell response in the face of an overly exuberant response largely involving  $T_H1$  and  $T_H17$  cells in Crohn's disease and  $T_H2$  cells in ulcerative colitis. Intestinal inflammation resulting from a failure to maintain this balance is exemplified by patients with immune dysregulation, polyendocrinopathy, enteropathy, X-linked (IPEX) syndrome or with WAS (also known as WASP) deficiency, who have deficient  $T_{reg}$ -cell function. Furthermore,  $T_H$ -cell polarization in IBD is unlikely to be a simple divergence between a few disparate T-cell fates, but rather to include diverse sub-programs that can be selectively activated by antigen-presenting cells, cytokine milieu, microbial factors and metabolic programs. This notion is supported by the findings that T cells expressing both IFN- $\gamma$  and IL-17 are detected during all stages of Crohn's disease, and it evolves our understanding of pro-inflammatory and anti-inflammatory cell types and pathways.

Recent studies indicate that  $T_{reg}$  cells and  $T_H17$  cells may arise from a common precursor, consistent with the observation that TGF- $\beta$  helps to direct differentiation of both subsets<sup>62</sup>. The generation of two subsets with opposing activities from a common precursor is reminiscent of differential responses of precursor cells along morphogen gradients during development, suggesting that similar morphogen gradients may work in the gut *in vivo*. In this regard, TGF- $\beta$  alone drives  $T_{reg}$ -cell differentiation, with retinoic acid exerting a synergistic effect; the differentiation and function of mucosal  $T_{reg}$  cells depends on the transcription factors BLIMP1 and IRF4 (ref. 63). Given the abundance of TGF- $\beta$  in intestinal tissues, this may contribute to baseline homeostasis, for example, by promoting  $T_{reg}$ -cell differentiation in naive lamina propria CD4<sup>+</sup> T cells. However, in conjunction with other signals, including cytokines, metabolites and microbial signals,  $T_H17$ -cell differentiation is promoted instead. Experiments demonstrating the crucial role of IL-23, IL-6 and IL-17 in the development of experimental colitis support a role for  $T_H17$  cells in disease propagation<sup>64</sup>. Illustrating some of the intercellular interactions that affect the  $T_H17$ - $T_{reg}$ -cell axis,  $\gamma\delta$  T cells can drive the  $T_H17$  program and contribute to experimental colitis, and are in turn suppressed by  $T_{reg}$  cells<sup>65,66</sup>. Furthermore,  $T_{reg}$  cells can support the development of  $T_H17$  cells by maintaining decreased levels of IL-2 in the local milieu<sup>67</sup>.

Transcriptional programs helmed by the  $T_{reg}$ - and  $T_H17$ -cell-lineage-defining transcription factors FOXP3 and ROR $\gamma$ t work together with a network of transcription factors, which can in turn respond to lineage-inducing cytokines and microbial factors. Transcription factors can mediate dichotomous functions depending on the cellular (and probably cytokine) context; for example, STAT3 drives  $T_H17$ -cell differentiation, but is also required for anti-inflammatory IL-10 signalling through distinct pathways, inducing repressors such as strawberry notch homologue 2 (SBNO2). The aryl hydrocarbon receptor (AHR) is a nuclear receptor that is essential for IL-22 production and also enhances IL-17 production, albeit to a lesser extent than ROR $\gamma$ t-driven pathways, illustrating how distinct transcription factors may drive separate functions in the same cell<sup>68</sup>. AHR responds to polycyclic hydrocarbons, suggesting that xenobiotic stimuli may modulate IL-22 and IL-17 production. Many of the genes required for  $T_{reg}$ - and  $T_H17$ -cell differentiation have been implicated in IBD (Fig. 3). *CCR6*, which lies in a locus associated with Crohn's disease, encodes a chemokine receptor that is important for lymphocyte homing to the gut as well as for the development of intestinal lymphoid follicles — areas important for the production of T-cell-independent IgA, which affects the microbiota composition of the host<sup>33</sup>. Thus, the gut may use TGF- $\beta$  pathways to poise T cells to carry out both pro- and anti-inflammatory programs depending on the local presence of cytokines and microbial products.

Illustrating the concept that hyper- or hypo-activation can affect the outcome of T-cell differentiation, defects in ITCH, a HECT-type E3 ubiquitin ligase involved in T-cell activation, lead to impaired  $T_{reg}$ -cell

polarization in mice and to autoimmunity in patients<sup>69</sup>. IBD-associated loci also contain other members of the ITCH pathway, including NDFIP1 and TNFAIP3. Defects in these proteins are associated with inappropriate T-cell activation, skewed  $T_H$ -cell polarization and pathological intestinal inflammation, consistent with the hypothesis that *NDFIP1* and *TNFAIP3* are disease-contributory genes<sup>70</sup>.

The soluble mediators secreted by  $T_{reg}$  cells are also released by other, FOXP3<sup>+</sup> regulatory T-cell subsets in the gut, such as T regulatory 1 ( $T_{r1}$ ) cells. IL-10 release can be induced by IL-27 in several subsets, including both pro-inflammatory  $T_H1$  and anti-inflammatory  $T_{r1}$  subsets. IL-27 is made by antigen-presenting cells, illustrating another homeostatic interaction between innate and adaptive immune cells<sup>71</sup>. Interestingly, GWAS implicate both *IL10* and *IL27* in IBD, suggesting that this may represent a central axis of immune regulation in the context of IBD.

In addition to the contribution of T-cell subsets, experimental evidence suggests that B-cell defects may contribute to the development of colitis in several ways, including impaired IgA production and antigen presentation, effects on early B-cell selection, and the perturbed production of pro- and anti-inflammatory mediators. Supporting the importance of IgA production in IBD, recent GWAS of selective IgA deficiency showed genes also implicated in IBD, namely *ORMDL3*, *REL* and *PTPN22* (ref. 72).

Immunoglobulins can also have immune modulatory activity, which is highlighted by IBD genetic studies. Peripheral B-cell tolerance can be maintained by signalling through the lectin CD22, which binds immunoglobulin bearing  $\alpha 2,6$ -linked sialic acid<sup>73</sup>. This epitope is generated in part through the action of sialic acid acetyltransferase (SIAE). Sequencing studies in small cohorts detected rare SIAE variants that disrupted enzyme function and secretion in patients with IBD and other autoimmune diseases<sup>73</sup>. In addition, sialylated IgG may signal through DC-SIGN (dendritic-cell-specific ICAM3-grabbing non-integrin) on myeloid cells, leading to increased expression of inhibitory Fc $\gamma$ RII $\beta$  (the gene encoding which lies in an ulcerative colitis-implicated locus) on macrophages. These data demonstrate pathways by which immunoglobulins can exert anti-inflammatory activities and highlight components that genetic studies suggest may be perturbed in IBD.

The functional relevance of anti-inflammatory B regulatory ( $B_{reg}$ ) cells has been demonstrated in several mouse models of inflammatory diseases, including colitis;  $B_{reg}$  cells from patients with SLE also show impaired function<sup>74</sup>.  $B_{reg}$  cells differentiate after stimulation in the context of either anti-CD40 antibody or TLR ligands, and secrete anti-inflammatory cytokines such as TGF- $\beta$  and IL-10. Defects in  $B_{reg}$ -cell development or function might lead to failure to upregulate IL-10, leading to attenuated suppression of CD4<sup>+</sup> T-cell production of IFN- $\gamma$  and TNF- $\alpha$ , consistent with a broader role of  $B_{reg}$  cells in autoimmunity and inflammatory disease.

Other cell types that help to regulate gut immunity include intra-epithelial lymphocytes, which comprise many subsets such as CD8 $\alpha\alpha$ <sup>+</sup>  $\gamma\delta$  T cells (which show both cytoprotective and cytolytic activities) and CD8 $\alpha\alpha$ <sup>+</sup>  $\alpha\beta$  T cells (which are thought to be regulatory cells that require TGF- $\beta$  for development)<sup>75</sup>. Enrichment analysis of expression profiles further suggests that a subset of IBD-implicated genes is expressed in natural killer T (NKT) cells, which can detect infection through microbial lipids presented by CD1d or through atypical endogenous lipids, such as isoglobotriaoisylceramide, which accumulate after microbial-TLR signalling<sup>76</sup>. Furthermore, NKT cells from patients with ulcerative colitis produce IL-13 and show enhanced cytotoxicity, further indicating that a perturbed NKT-cell response to as yet unidentified bacterial ligands may be pro-colitogenic.

Genetic studies may offer insight into how this balance is disturbed, leading to pathological inflammation. Interestingly, perturbation by blocking cytotoxic T-lymphocyte antigen 4, an important inhibitory molecule expressed on activated T cells, commonly resulted in colitis in patients, again suggesting the poised state of activation of T cells in the gut<sup>77</sup>.



## Mucosal ecology and immune responses in disease

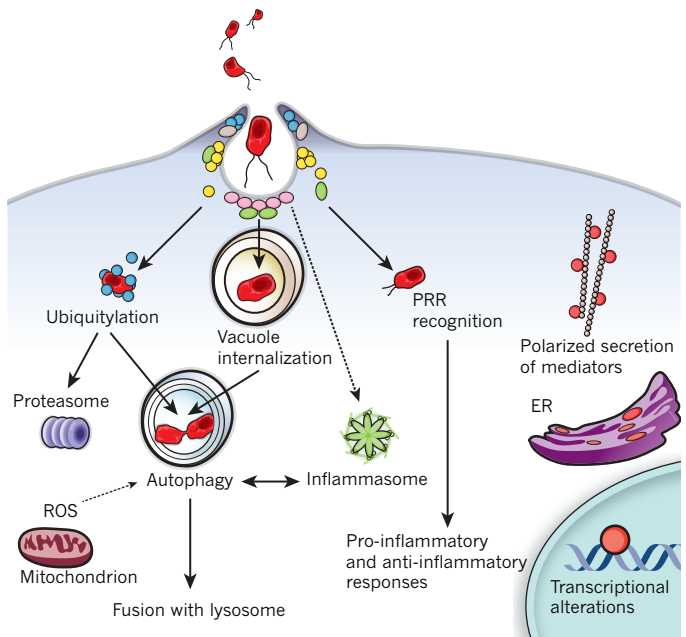
The gut microflora is a community that has co-evolved with the host and confers beneficial effects, including helping to metabolize nutrients, modulate immune responses and defend against pathogens. However, dysregulation of normal co-evolved homeostatic relationships between gut bacteria and host immune responses can lead to intestinal inflammation. Indeed, accumulating evidence suggests that luminal flora is a requisite, perhaps even a central factor in the development of IBD.

Efforts to correlate changes in enteric microbial communities with disease are complicated by the great interindividual variation, such that even monozygotic twins may share only 40% of faecal phylotypes<sup>78</sup>. However, clustering the abundance of genes in certain categories in a species-independent fashion shows high interindividual similarity, suggesting that the microbiome can be perceived as a conserved functional entity<sup>79</sup>. Differences in the abundance of both bacterial species and functional gene categories (such as bacterial motility, sugar and iron metabolism) can differentiate patients with IBD from healthy individuals, demonstrating IBD-related changes in gut microbial ecology<sup>80</sup> (C. Huttenhower, personal communication).

Even within an individual, intestinal microbial communities are dynamic and influenced by host factors, dietary effects and the microbes themselves. Many of the specific examples have emerged from studies in mice. Illustrating how microbial communities can be affected by host responses, infection-induced inflammation results in an oxidative metabolic shift in *Salmonella enterica* serovar Typhimurium (*S. Typhimurium*), which the bacterium uses in conjunction with host-derived ROS to create a growth advantage over fermenting bacteria<sup>81</sup>. Microbe–host relationships are tightly interrelated, such that host factors can induce functional changes in the microflora that, in return, affect host biology. TLRs recognize microbial motifs and have a crucial role in determining mucosal susceptibility to injury and repair responses. Impaired TLR signalling due to *Myd88* deficiency in non-obese diabetic (NOD) mice induces changes in microbiota community structure that protect against diabetes<sup>82</sup>. Conversely, impaired innate immune function in *T-bet*<sup>−/−</sup> *Rag1*<sup>−/−</sup> (also known as *Tbx21*<sup>−/−</sup> *Rag1*<sup>−/−</sup>) mice and *Tlr5*<sup>−/−</sup> mice leads to the generation of a pathogenic microbiota that causes colitis and metabolic syndrome, respectively, even in genetically normal hosts<sup>83,84</sup>. Similarly, the gut microbiota induces a dynamic IgA response, with qualitative and/or quantitative defects in IgA production resulting in impaired control of the microbial communities<sup>85,86</sup>. Deficiencies in activation-induced cytidine deaminase — an enzyme essential for somatic hypermutation and class-switching recombination during B-cell maturation — result in IgA deficiency, specific expansion of the anaerobic flora and segmented filamentous bacteria (SFB), and overstimulation of the mucosal immune system, with hyperplasia of mucosal lymphoid structures such as Peyer's patches and lymphoid follicles. Similar observations in mice with selectively impaired somatic hypermutation point to the importance of affinity maturation in generating diversity in the IgA repertoire to control the intestinal microbial burden<sup>86</sup>.

Mice fostered on milk lacking sialyl(α2,3)lactose develop a distinct microbiota that confers transmissible resistance to DSS colitis, providing an example of dietary effects on the gut microbiota<sup>87</sup>. Dietary glycans can also be incorporated onto host cell membranes and can act as receptors for bacterial toxins. These findings demonstrate that host factors, both transient and genetic, can act together with dietary factors to modulate microbiota community structure and/or function, sometimes indelibly, in IBD-relevant ways.

The intestinal mucosa can monitor microbial ligands using pattern recognition receptors (PRRs), and microbial metabolites using G-protein-coupled receptors (GPCRs) and solute carriers. Short-chain fatty acids (SCFAs), generated by some microflora constituents and decreased in ulcerative colitis, can signal through the receptor GPR43 in neutrophils, with notable proresolving effects on inflammation<sup>88</sup>. Other examples of GPCRs with immune modulatory activity include GPR120,



**Figure 5 | Intracellular defence programs in microbial recognition.** Host cells have evolved processes by which they restrict the availability of intracellular permissive niches to microbes. Microbial recognition by PRRs, such as NOD proteins and TLRs, activates key immediate host programs, leading to polarized secretion of pro-inflammatory mediators (directed to either the luminal or basolateral surface). Bacteria can either be maintained in subcellular compartments such as microbe-containing vacuoles, or escape into the cytoplasm, where they can be ubiquitylated and targeted for degradation. Both subsets can be targeted by the autophagy pathway, which is also regulated by other host defence mechanisms such as oxidative stress and inflammasome activation.

GPR65 (also known as TDAG8) and GPR35, which can be activated by ω-3 fatty acids, extracellular protons and kynurenic acid, respectively, with anti-inflammatory effects<sup>89</sup>. Of interest, kynurenic acid and other anti-inflammatory kynurenines are generated by the catabolism of tryptophan. Host levels of tryptophan are affected by the microbiota, suggesting how microbes can modulate the host immune response by metabolic effects. Other microbial metabolites can be transported into host cells. For example, the solute carrier SLC22A5 transports a quorum-sensing molecule from *Bacillus subtilis*, conferring resistance to oxidative stress *in vitro*<sup>90</sup>. The proton-coupled histidine/peptide cotransporter SLC15A4 is required for TLR7 and TLR9 signalling in plasmacytoid dendritic cells<sup>91</sup>. Furthermore, studies in *Slc15a4*<sup>−/−</sup> dendritic cells suggest that SLC15A4 contributes to TLR9 signalling by regulating endosomal histidine levels, and to NOD1 signalling by cytosolic delivery of NOD1 ligands such as the tripeptide motif L-Ala-γ-D-Glu-meso-diaminopimelic acid (Tri<sub>DAB</sub>). *Slc15a4* deficiency ameliorates susceptibility to DSS colitis in mice<sup>91</sup>. These findings and the presence of genes such as *SLC22A5*, *GPR35* and *GPR65* in IBD-risk loci suggest that PRRs, GPCRs and solute carriers help to maintain microbe–host relationships and intestinal homeostasis by transducing signals from microbial ligands and metabolites, which can in turn have immune modulatory effects on the host.

Microbial signals also shape innate and adaptive immune responses. Germ-free animals have underdeveloped Peyer's patches, as well as fewer IgA-producing plasma cells and lamina propria CD4<sup>+</sup> cells, illustrating the role of the microbiota in generating a mature mucosal adaptive immune response. The constituents of the microbiota can have important protective roles; for example, the impaired epithelial injury response in *Myd88*-deficient mice highlights the role of microbial stimulation in epithelial restitution<sup>92</sup>. Similarly, in mice, commensal bacteria activate expression of the transcription factor NFIL3, inhibit

*IL12b* expression and protect against colitis, and CD14<sup>+</sup> lamina propria cells from patients with IBD express less NFIL3 than healthy controls<sup>93</sup>. The microbiota also acts on the epithelium together with adaptive immune signals, inducing epithelial secretion of IL-25, which represses ILC secretion of IL-22, and thus IL-22-induced AMPs. This equilibrium in the healthy mucosa is abrogated by epithelial insults, leading to increased IL-22 and activation of antimicrobial programs<sup>40</sup>.

Microbial populations and ligands can have pro-inflammatory or anti-inflammatory effects. In mice, SFB promote T<sub>H</sub>17 differentiation and IgA production, whereas *Clostridium* clusters IV and XIVa and parasite-secreted proteins such as *Heligmosomoides polygyrus* excretory-secretory antigen promote T<sub>reg</sub>-cell differentiation<sup>94–97</sup>. Interestingly, patients with IBD show reduced representation of *Clostridium* clusters IV and XIVa, indicating one way in which anti-inflammatory T<sub>reg</sub>-cell effects might be diminished, leading to a predisposition to inflammation<sup>98</sup>. The common constituent of normal human microflora *Bacteroides fragilis* produces polysaccharide A, which suppresses IL-17 production and promotes the activity of IL-10-producing CD4<sup>+</sup> T cells in mice<sup>99</sup>. The effects of microbial ligands and metabolites on adaptive immune function are exemplified by bacterial DNA signalling through TLR9 to limit T<sub>reg</sub>-cell differentiation and promote intestinal immune responses to oral infection, and by bacterial ATP promoting T<sub>H</sub>17 differentiation. Thus, the microflora shapes development and function of the mucosal immune system in a tightly correlated manner. Immune stimulatory effects of the microbiota are important to promote an effective response against potential pathogens, although dysregulated interactions, which might arise from perturbations in host, microbial or environmental factors, could lead to a loss of tolerance and promote intestinal inflammation.

Most of the observations detailing the mechanisms of microbe–host interactions have been made in mice, and correlations in humans remain to be defined. Microbes associated with human IBD include *Faecalibacterium prausnitzii*, adherent-invasive *E. coli*, invasive *Fusobacterium nucleatum* and mucolytic bacteria such as *Ruminococcus gnavus* and *Ruminococcus torques*. Reduced levels of *F. prausnitzii* in resected ileal mucosa from patients with Crohn's disease are associated with increased risk of endoscopic recurrence; *F. prausnitzii* stimulates IL-10 production in peripheral blood mononuclear cells, which may account at least in part for this protective effect<sup>100</sup>. Recent studies suggest that adherent-invasive *E. coli* exploits host defects in phagocytosis and autophagy arising from Crohn's-disease-related polymorphisms to promote chronic inflammation in the susceptible host<sup>16</sup>. Patients with IBD have a compromised mucus layer and an epithelial surface that is densely coated with bacteria; the abundant presence of *Ruminococcus* strains in IBD mucosa raises the possibility that such microbes may contribute to the barrier defect observed in IBD, although whether their presence is causal or correlative remains unclear.

These findings show that the composition of the microbiota and its interaction with the host are emerging as underappreciated sources of gene–environment interactions and are crucial to understanding the context of IBD. For example, alterations in the microflora community structure, as might occur in the context of antibiotic therapy or infectious colitis, can promote the development of IBD or trigger disease flares in patients with IBD. Identifying the factors that shape microbial community structure and function within an individual and that influence its restoration after perturbations will be key to understanding IBD pathogenesis. Obtaining such knowledge will require identifying associations between microbiome and human genetic studies at the very least.

## Future perspectives

GWAS and next-generation sequencing technologies have provided insight into genetic definitions of host susceptibility. GWAS have unequivocally identified numerous genomic regions containing IBD-risk factors, showing several features of the genetic architecture of Crohn's disease and ulcerative colitis. First, IBD risk involves

multigenic contributions, each with a relatively modest effect size. Second, genetic contributions to ulcerative colitis and Crohn's disease overlap, suggesting shared mechanistic features. Third, within Crohn's disease and ulcerative colitis, different clusters of risk loci are emerging, suggesting that these disease processes may comprise distinct pathological subsets beyond Crohn's disease versus ulcerative colitis. Accordingly, there is a need to define clinically relevant parameters that might help to classify Crohn's disease and ulcerative colitis further, including early-onset disease, stricturing disease, slow progressors, frequency of flares and response to therapeutics. Furthermore, given the importance of environmental factors in IBD risk, studies aimed at defining contributory environmental factors are greatly needed. Relevant approaches might include establishing prospective inception cohorts or following healthy, high-risk individuals, such as those with an affected first-degree relative.

An important adjunctive approach to GWAS is identifying rare variants, which frequently show larger effect sizes. The search for rare variants will help to prioritize the probable causal gene(s) within a locus (or loci) for experimental validation, identify disease-relevant pathways, and possibly identify domains important for protein function by leveraging natural mutations as a large forward genetic screen. Identifying and validating causal genes and assembling them into molecular pathways and cellular networks will require the use of patient samples and will considerably empower clinically relevant hypotheses. Given the diverse mechanisms that seem to participate in IBD, it will also become increasingly important to associate and stratify 'omic' measurements of RNA, protein, small molecules, chemical DNA modifications and gut microbiota according to patient genotypes.

There is a clear need to generate quantitative and qualitative expression maps of allelic variants. This notion is reinforced by the many polymorphisms implicating gene deserts, which probably contain regulatory elements. Furthermore, alternative splicing is a major contributor to the diversity of our transcriptome and its relevance to IBD has already been demonstrated by findings in *IL23R*.

The gut has many tiers of defence against incursion by luminal microbes, including the epithelial barrier, and the innate and adaptive immune responses. These components are all tightly interrelated, and disease requires breakdown at several checkpoints. Generating models to systematically analyse the defects arising from genetic variants associated with IBD is crucial. However, these variants may show the disease-relevant defect under select conditions, such as high bacterial load found in the colon, and the accompanying cytokine milieu.

Viral infections are common, and key studies highlight their potential to exert important immune modulatory effects. Acute and/or chronic viral infections could interact with host-susceptibility factors in a manner that leaves either the cell or the cellular milieu poised to promote pathological intestinal inflammation after subsequent triggering events. Notably, these studies highlight the need to characterize all microbial constituents (viral, fungal, parasitic and bacterial) in the context of IBD. Other tools need to be developed to study the microbiota at the level of species, geographical location, genetic variations, transcriptional dynamics, as well as changes to proteins and metabolites. Indeed, bacterial metabolites are principal mediators of interactions between microbial species, as well as between microbe and host, as exemplified by SCFAs. Studies to identify bioactive metabolites and other small molecules may thus have diagnostic and therapeutic potential.

An important goal is to combine these various facets to understand how genetic traits are integrated and propagated through physiological networks in the context of interactions with other genes, cells, microbes and environmental stimuli to control intestinal homeostasis. Genetic studies are already used to predict sensitivity to IBD therapies such as 6-mercaptopurine and may also be useful in predicting responses to biological therapies. Combining the different aspects of IBD pathophysiology may allow us to develop a more holistic understanding of the disease, thus promoting advances in diagnostics and therapy. ■



1. Franke, A. *et al.* Genome-wide meta-analysis increases to 71 the number of confirmed Crohn's disease susceptibility loci. *Nature Genet.* **42**, 1118–1125 (2010).  
**This study identifies 39 new loci containing genes in previously identified Crohn's-disease-relevant pathways, including IL10, IL27 and TYK2, and reports the important genetic overlap with loci implicated in other immune-related diseases.**
2. Anderson, C. A. *et al.* Meta-analysis identifies 29 additional ulcerative colitis risk loci, increasing the number of confirmed associations to 47. *Nature Genet.* **43**, 246–252 (2011).  
**This study identifies 29 new loci containing genes that reinforce the contribution of epithelial barrier function, intracellular defence and cytokine-dependent signalling in ulcerative colitis, and it also reports genetic overlap with Crohn's disease.**
3. Spehlmann, M. E. *et al.* Epidemiology of inflammatory bowel disease in a German twin cohort: results of a nationwide study. *Inflamm. Bowel Dis.* **14**, 968–976 (2008).
4. Yazdanyar, S., Kamstrup, P. R., Tybjaerg-Hansen, A. & Nordestgaard, B. G. Penetrance of *NOD2/CARD15* genetic variants in the general population. *Can. Med. Assoc. J.* **182**, 661–665 (2010).
5. Monick, M. M. *et al.* Identification of an autophagy defect in smokers' alveolar macrophages. *J. Immunol.* **185**, 5425–5435 (2010).
6. Momozawa, Y. *et al.* Resequencing of positional candidates identifies low frequency *IL23R* coding variants protecting against inflammatory bowel disease. *Nature Genet.* **43**, 43–47 (2010).
7. Wang, K. *et al.* Comparative genetic analysis of inflammatory bowel disease and type 1 diabetes implicates multiple loci with opposite effects. *Hum. Mol. Genet.* **19**, 2059–2067 (2010).
8. Janse, M. *et al.* Three ulcerative colitis susceptibility loci are associated with primary sclerosing cholangitis and indicate a role for *IL2*, *REL* and *CARD9*. *Hepatology* **53**, 1977–1985 (2011).
9. Zhang, F.-R. *et al.* Genomewide association study of leprosy. *N. Engl. J. Med.* **361**, 2609–2618 (2009).
10. Waisberg, M. *et al.* Genetic susceptibility to systemic lupus erythematosus protects against cerebral malaria in mice. *Proc. Natl Acad. Sci. USA* **108**, 1122–1127 (2011).
11. Genovese, G. *et al.* Association of trypanolytic ApoL1 variants with kidney disease in African Americans. *Science* **329**, 841–845 (2010).
12. Glocker, E.-O. *et al.* Inflammatory bowel disease and mutations affecting the interleukin-10 receptor. *N. Engl. J. Med.* **361**, 2033–2045 (2009).
13. Di Meglio, P. *et al.* The *IL23R* R381Q gene variant protects against immune-mediated diseases by impairing IL-23-induced Th17 effector response in humans. *PLoS ONE* **6**, e17160 (2011).
14. Yu, R. Y. & Gallagher, G. A naturally occurring, soluble antagonist of human IL-23 inhibits the development and *in vitro* function of human Th17 cells. *J. Immunol.* **185**, 7302–7308 (2010).
15. Rosenstiel, P. *et al.* A short isoform of *NOD2/CARD15*, *NOD2-S*, is an endogenous inhibitor of *NOD2*/receptor-interacting protein kinase 2-induced signaling pathways. *Proc. Natl Acad. Sci. USA* **103**, 3280–3285 (2006).
16. Brest, P. *et al.* A synonymous variant in *IRGM* alters a binding site for miR-196 and causes deregulation of *IRGM*-dependent xenophagy in Crohn's disease. *Nature Genet.* **43**, 242–245 (2011).
17. Khalil, A. M. *et al.* Many human large intergenic noncoding RNAs associate with chromatin-modifying complexes and affect gene expression. *Proc. Natl Acad. Sci. USA* **106**, 11667–11672 (2009).
18. Muise, A. M. *et al.* Polymorphisms in E-cadherin (*CDH1*) result in a mis-localised cytoplasmic protein that is associated with Crohn's disease. *Gut* **58**, 1121–1127 (2009).
19. Sabath, E. *et al.* Ga12 regulates protein interactions within the MDCK cell tight junction and inhibits tight-junction assembly. *J. Cell Sci.* **121**, 814–824 (2008).
20. Scharl, M. *et al.* Protection of epithelial barrier function by the Crohn's disease associated gene protein tyrosine phosphatase N2. *Gastroenterology* **137**, 2030–2040.e5 (2009).
21. Hassan, S. W. *et al.* Increased susceptibility to dextran sulfate sodium induced colitis in the T cell protein tyrosine phosphatase heterozygous mouse. *PLoS ONE* **5**, e8868 (2010).
22. Darsigny, M. *et al.* Loss of hepatocyte-nuclear-factor-4a affects colonic ion transport and causes chronic inflammation resembling inflammatory bowel disease in mice. *PLoS ONE* **4**, e7609 (2009).
23. Cattin, A.-L. *et al.* Hepatocyte nuclear factor 4a, a key factor for homeostasis, cell architecture, and barrier function of the adult intestinal epithelium. *Mol. Cell. Biol.* **29**, 6294–6308 (2009).
24. Pabst, O., Zweigerdt, R. & Arnold, H. H. Targeted disruption of the homeobox transcription factor *Nkx2-3* in mice results in postnatal lethality and abnormal development of small intestine and spleen. *Development* **126**, 2215–2225 (1999).
25. Pickert, G. *et al.* STAT3 links IL-22 signaling in intestinal epithelial cells to mucosal wound healing. *J. Exp. Med.* **206**, 1465–1472 (2009).
26. Fu, J. *et al.* Loss of intestinal core 1-derived O-glycans causes spontaneous colitis in mice. *J. Clin. Invest.* **121**, 1657–1666 (2011).
27. Cadwell, K. *et al.* A key role for autophagy and the autophagy gene *Atg16l1* in mouse and human intestinal Paneth cells. *Nature* **456**, 259–263 (2008).
28. Biswas, A. *et al.* Induction and rescue of *Nod2*-dependent Th1-driven granulomatous inflammation of the ileum. *Proc. Natl Acad. Sci. USA* **107**, 14739–14744 (2010).
29. Kaser, A. *et al.* XBP1 links ER stress to intestinal inflammation and confers genetic risk for human inflammatory bowel disease. *Cell* **134**, 743–756 (2008).  
**This study illustrates the role of Paneth and goblet cells in intestinal homeostasis and how the control of ER stress is crucial for their maintenance, such that XBP1 deficiency leads to Paneth cell loss, reduced goblet cell numbers and increased susceptibility to experimental colitis.**
30. Brandl, K. *et al.* Enhanced sensitivity to DSS colitis caused by a hypomorphic *Mbtps1* mutation disrupting the ATF6-driven unfolded protein response. *Proc. Natl Acad. Sci. USA* **106**, 3300–3305 (2009).
31. Goodall, J. C. *et al.* Endoplasmic reticulum stress-induced transcription factor, CHOP, is crucial for dendritic cell IL-23 expression. *Proc. Natl Acad. Sci. USA* **107**, 17698–17703 (2010).
32. Campbell, E. L. *et al.* Resolvin E1-induced intestinal alkaline phosphatase promotes resolution of inflammation through LPS detoxification. *Proc. Natl Acad. Sci. USA* **107**, 14298–14303 (2010).
33. Rescigno, M. Intestinal dendritic cells. *Adv. Immunol.* **107**, 109–138 (2010).
34. Blasius, A. L. *et al.* Slc15a4, AP-3, and Hermansky-Pudlak syndrome proteins are required for Toll-like receptor signaling in plasmacytoid dendritic cells. *Proc. Natl Acad. Sci. USA* **107**, 19973–19978 (2010).
35. Smith, A. M. *et al.* Disordered macrophage cytokine secretion underlies impaired acute inflammation and bacterial clearance in Crohn's disease. *J. Exp. Med.* **206**, 1883–1897 (2009).
36. Amit, I. *et al.* Unbiased reconstruction of a mammalian transcriptional network mediating pathogen responses. *Science* **326**, 257–263 (2009).  
**This pioneering study uses computational analyses of time-course gene-expression profiles to characterize dendritic cell programs and predict signalling circuits engaged after stimulation by microbe-associated molecules, and it experimentally validates key regulators.**
37. Siddiqui, K. R. R., Laffont, S. & Powrie, F. E-cadherin marks a subset of inflammatory dendritic cells that promote T cell-mediated colitis. *Immunity* **32**, 557–567 (2010).
38. Uematsu, S. *et al.* Detection of pathogenic intestinal bacteria by Toll-like receptor 5 on intestinal CD11c<sup>+</sup> lamina propria cells. *Nature Immunol.* **7**, 868–874 (2006).
39. Swirski, F. K. *et al.* Identification of splenic reservoir monocytes and their deployment to inflammatory sites. *Science* **325**, 612–616 (2009).
40. Sawa, S. *et al.* RORγ<sup>+</sup> innate lymphoid cells regulate intestinal homeostasis by integrating negative signals from the symbiotic microbiota. *Nature Immunol.* **12**, 320–326 (2011).
41. Cella, M. *et al.* A human natural killer cell subset provides an innate source of IL-22 for mucosal immunity. *Nature* **457**, 722–725 (2009).  
**This study describes mucosal IL-22-secreting ILCs in humans and mice, capturing many key properties of these cells, including their ability to respond to IL-23 stimulation and promote epithelial restitutive effects.**
42. Takayama, T. *et al.* Imbalance of Nkp44<sup>+</sup>Nkp46<sup>+</sup> and Nkp44<sup>+</sup>Nkp46<sup>+</sup> natural killer cells in the intestinal mucosa of patients with Crohn's disease. *Gastroenterology* **139**, 882–892.e3 (2010).
43. Shaw, M. H., Kamada, N., Warner, N., Kim, Y. G. & Nunez, G. The ever-expanding function of *NOD2*: autophagy, viral recognition, and T cell activation. *Trends Immunol.* **32**, 73–79 (2011).
44. Cooney, R. *et al.* *NOD2* stimulation induces autophagy in dendritic cells influencing bacterial handling and antigen presentation. *Nature Med.* **16**, 90–97 (2010).
45. Travassos, L. H. *et al.* *Nod1* and *Nod2* direct autophagy by recruiting ATG16L1 to the plasma membrane at the site of bacterial entry. *Nature Immunol.* **11**, 55–62 (2010).
46. Sabbah, A. *et al.* Activation of innate immune antiviral responses by *Nod2*. *Nature Immunol.* **10**, 1073–1080 (2009).
47. Zaki, M. H. *et al.* The NLRP3 inflammasome protects against loss of epithelial integrity and mortality during experimental colitis. *Immunity* **32**, 379–391 (2010).
48. Hsu, Y.-M. S. *et al.* The adaptor protein *CARD9* is required for innate immune responses to intracellular pathogens. *Nature Immunol.* **8**, 198–205 (2007).
49. Poeck, H. *et al.* Recognition of RNA virus by RIG-I results in activation of *CARD9* and inflammasome signaling for interleukin 1β production. *Nature Immunol.* **11**, 63–69 (2010).
50. Gringhuis, S. I. *et al.* Selective C-Rel activation via Malt1 controls anti-fungal T<sub>H</sub>-17 immunity by dectin-1 and dectin-2. *PLoS Pathogens* **7**, e1001259 (2011).
51. Schroeder, B. O. *et al.* Reduction of disulphide bonds unmasks potent antimicrobial activity of human β-defensin 1. *Nature* **469**, 419–423 (2011).
52. Dolowisch, T. *et al.* Potentiation of epithelial innate host responses by intercellular communication. *PLoS Pathogens* **6**, e1001194 (2010).
53. Gardet, A. *et al.* LRRK2 is involved in the IFN-γ response and host response to pathogens. *J. Immunol.* **185**, 5577–5585 (2010).
54. Wu, W., Hsu, Y.-M. S., Bi, L., Songyang, Z. & Lin, X. *CARD9* facilitates microbe-elicited production of reactive oxygen species by regulating the LyGDI-Rac1 complex. *Nature Immunol.* **10**, 1208–1214 (2009).
55. Efimova, O., Szankasi, P. & Kelley, T. W. Ncf1 (p47phox) is essential for direct regulatory T cell mediated suppression of CD4<sup>+</sup> effector T cells. *PLoS ONE* **6**, e16013 (2011).
56. Kraaij, M. D. *et al.* Induction of regulatory T cells by macrophages is dependent on production of reactive oxygen species. *Proc. Natl Acad. Sci. USA* **107**, 17686–17691 (2010).
57. Rioux, J. D. *et al.* Genome-wide association study identifies new susceptibility loci for Crohn disease and implicates autophagy in disease pathogenesis. *Nature Genet.* **39**, 596–604 (2007).
58. Hampe, J. *et al.* A genome-wide association scan of nonsynonymous SNPs identifies a susceptibility variant for Crohn disease in *ATG16L1*. *Nature Genet.* **39**, 207–211 (2007).

59. McCarroll, S. A. *et al.* Deletion polymorphism upstream of *IRGM* associated with altered *IRGM* expression and Crohn's disease. *Nature Genet.* **40**, 1107–1112 (2008).
60. Kuballa, P., Huett, A., Rioux, J. D., Daly, M. J. & Xavier, R. J. Impaired autophagy of an intracellular pathogen induced by a Crohn's disease associated *ATG16L1* variant. *PLoS ONE* **3**, e3391 (2008).
61. Cadwell, K. *et al.* Virus-plus-susceptibility gene interaction determines Crohn's disease gene *Atg16L1* phenotypes in intestine. *Cell* **141**, 1135–1145 (2010).  
**This seminal paper reports that a Crohn's disease susceptibility gene requires interaction with environmental (microbial) cues to manifest a disease phenotype.**
62. Zhou, L. *et al.* TGF- $\beta$ -induced Foxp3 inhibits T<sub>H</sub>17 cell differentiation by antagonizing ROR $\gamma$ t function. *Nature* **453**, 236–240 (2008).
63. Cretney, E. *et al.* The transcription factors Blimp-1 and IRF4 jointly control the differentiation and function of effector regulatory T cells. *Nature Immunol.* **12**, 304–311 (2011).
64. Ghoreschi, K. *et al.* Generation of pathogenic T<sub>H</sub>17 cells in the absence of TGF- $\beta$  signalling. *Nature* **467**, 967–971 (2010).
65. Do, J. S., Visperas, A., Dong, C., Baldwin, W. M. III & Min, B. Generation of colitogenic Th17 CD4 T cells is enhanced by IL-17\*  $\gamma\delta$  T cells. *J. Immunol.* **186**, 4546–4550 (2011).
66. Park, S.-G. *et al.* T regulatory cells maintain intestinal homeostasis by suppressing  $\gamma\delta$  T cells. *Immunity* **33**, 791–803 (2010).
67. Chen, Y. *et al.* Foxp3\* regulatory T cells promote T helper 17 cell development *in vivo* through regulation of interleukin-2. *Immunity* **34**, 409–421 (2011).
68. Trifari, S., Kaplan, C. D., Tran, E. H., Crellin, N. K. & Spits, H. Identification of a human helper T cell population that has abundant production of interleukin 22 and is distinct from T<sub>H</sub>-17, T<sub>H</sub>1 and T<sub>H</sub>2 cells. *Nature Immunol.* **10**, 864–871 (2009).
69. Lohr, N. J. *et al.* Human ITCH E3 ubiquitin ligase deficiency causes syndromic multisystem autoimmune disease. *Am. J. Hum. Genet.* **86**, 447–453 (2010).
70. Ramon, H. E. *et al.* The ubiquitin ligase adaptor Ndfip1 regulates T cell-mediated gastrointestinal inflammation and inflammatory bowel disease susceptibility. *Mucosal Immunol.* **4**, 314–324 (2010).
71. Cox, J. H. *et al.* IL-27 promotes T cell-dependent colitis through multiple mechanisms. *J. Exp. Med.* **208**, 115–123 (2011).
72. Ferreira, R. C. *et al.* Association of *IFIH1* and other autoimmunity risk alleles with selective IgA deficiency. *Nature Genet.* **42**, 777–780 (2010).
73. Surolia, I. *et al.* Functionally defective germline variants of sialic acid acetyltransferase in autoimmunity. *Nature* **466**, 243–247 (2010).
74. Blair, P. A. *et al.* CD19\*CD24<sup>hi</sup>CD38<sup>hi</sup> B cells exhibit regulatory capacity in healthy individuals but are functionally impaired in systemic lupus erythematosus patients. *Immunity* **32**, 129–140 (2010).
75. Konkel, J. E. *et al.* Control of the development of CD8 $\alpha^+$  intestinal intraepithelial lymphocytes by TGF- $\beta$ . *Nature Immunol.* **12**, 312–319 (2011).
76. Darmon, A. *et al.* Lysosomal  $\alpha$ -galactosidase controls the generation of self lipid antigens for natural killer T cells. *Immunity* **33**, 216–228 (2010).
77. Boasberg, P., Hamid, O. & O'Day, S. Ipilimumab: unleashing the power of the immune system through CTLA-4 blockade. *Semin. Oncol.* **37**, 440–449 (2010).
78. Turnbaugh, P. J. *et al.* Organismal, genetic, and transcriptional variation in the deeply sequenced gut microbiomes of identical twins. *Proc. Natl Acad. Sci. USA* **107**, 7503–7508 (2010).
79. Turnbaugh, P. J. *et al.* A core gut microbiome in obese and lean twins. *Nature* **457**, 480–484 (2009).
80. Qin, J. *et al.* A human gut microbial gene catalogue established by metagenomic sequencing. *Nature* **464**, 59–65 (2010).
81. Winter, S. E. *et al.* Gut inflammation provides a respiratory electron acceptor for *Salmonella*. *Nature* **467**, 426–429 (2010).  
**This paper shows how *S. Typhimurium* can use ROS generated during inflammation to form an electron acceptor that confers a growth advantage, illustrating how gut environment can affect microbial communities by providing a selective advantage to some bacterial species.**
82. Wen, L. *et al.* Innate immunity and intestinal microbiota in the development of type 1 diabetes. *Nature* **455**, 1109–1113 (2008).
83. Garrett, W. S. *et al.* Communicable ulcerative colitis induced by T-bet deficiency in the innate immune system. *Cell* **131**, 33–45 (2007).
84. Vijay-Kumar, M. *et al.* Metabolic syndrome and altered gut microbiota in mice lacking Toll-like receptor 5. *Science* **328**, 228–231 (2010).
85. Hapfelmeier, S. *et al.* Reversible microbial colonization of germ-free mice reveals the dynamics of IgA immune responses. *Science* **328**, 1705–1709 (2010).
86. Wei, M. *et al.* Mice carrying a knock-in mutation of *Aicda* resulting in a defect in somatic hypermutation have impaired gut homeostasis and compromised mucosal defense. *Nature Immunol.* **12**, 264–270 (2011).
87. Fuhrer, A. *et al.* Milk sialyllactose influences colitis in mice through selective intestinal bacterial colonization. *J. Exp. Med.* **207**, 2843–2854 (2010).
88. Maslowski, K. M. *et al.* Regulation of inflammatory responses by gut microbiota and chemoattractant receptor GPR43. *Nature* **461**, 1282–1286 (2009).  
**This paper highlights how commensal populations can modulate the host immune response, identifying a bacterial metabolic product (SCFA) and its receptor (GPR43), and shows that this pathway can ameliorate experimental colitis.**
89. Mogi, C. *et al.* Involvement of proton-sensing TDAG8 in extracellular acidification-induced inhibition of proinflammatory cytokine production in peritoneal macrophages. *J. Immunol.* **182**, 3243–3251 (2009).
90. Fujiya, M. *et al.* The *Bacillus subtilis* quorum-sensing molecule CSF contributes to intestinal homeostasis via OCTN2, a host cell membrane transporter. *Cell Host Microbe* **1**, 299–308 (2007).
91. Sasawata, S. *et al.* The solute carrier family 15A4 regulates TLR9 and NOD1 functions in the innate immune system and promotes colitis in mice. *Gastroenterology* **140**, 1513–1525 (2011).
92. Brandl, K. *et al.* MyD88 signaling in nonhematopoietic cells protects mice against induced colitis by regulating specific EGF receptor ligands. *Proc. Natl Acad. Sci. USA* **107**, 19967–19972 (2010).
93. Kobayashi, T. *et al.* NFIL3 is a regulator of IL-12 p40 in macrophages and mucosal immunity. *J. Immunol.* **186**, 4649–4655 (2011).
94. Ivanov, I. I. *et al.* Induction of intestinal Th17 cells by segmented filamentous bacteria. *Cell* **139**, 485–498 (2009).  
**This important study shows how particular components of the gut microbiota can modulate the adaptive immune response, with SFB promoting differentiation of pro-inflammatory T<sub>H</sub>17 cells.**
95. Gaboriau-Routhiau, V. *et al.* The key role of segmented filamentous bacteria in the coordinated maturation of gut helper T cell responses. *Immunity* **31**, 677–689 (2009).
96. Atarashi, K. *et al.* Induction of colonic regulatory T cells by indigenous *Clostridium* species. *Science* **331**, 337–341 (2010).  
**This study demonstrates that *Clostridium* clusters IV and XIVa affect the adaptive immune system by inducing a TGF- $\beta$ -rich milieu and promoting differentiation of anti-inflammatory T<sub>reg</sub> cells.**
97. Grainger, J. R. *et al.* Helminth secretions induce *de novo* T cell Foxp3 expression and regulatory function through the TGF- $\beta$  pathway. *J. Exp. Med.* **207**, 2331–2341 (2010).
98. Frank, D. N. *et al.* Molecular-phylogenetic characterization of microbial community imbalances in human inflammatory bowel diseases. *Proc. Natl Acad. Sci. USA* **104**, 13780–13785 (2007).
99. Mazmanian, S. K., Round, J. L. & Kasper, D. L. A microbial symbiosis factor prevents intestinal inflammatory disease. *Nature* **453**, 620–625 (2008).
100. Sokol, H. *et al.* *Faecalibacterium prausnitzii* is an anti-inflammatory commensal bacterium identified by gut microbiota analysis of Crohn disease patients. *Proc. Natl Acad. Sci. USA* **105**, 16731–16736 (2008).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** R.J.X., A.G. and B.K. are supported by grants from the National Institutes of Health and the Crohn's and Colitis Foundation of America. We apologize to those whose work is not cited owing to space constraints.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). Readers are welcome to comment on the online version of this article at [www.nature.com/nature](http://www.nature.com/nature). The authors declare no competing financial interests. Correspondence should be addressed to R.J.X. ([xavier@molbio.mgh.harvard.edu](mailto:xavier@molbio.mgh.harvard.edu)).



# Microenvironmental regulation of stem cells in intestinal homeostasis and cancer

Jan Paul Medema<sup>1</sup> & Louis Vermeulen<sup>1</sup>

**The identification of intestinal stem cells as well as their malignant counterparts, colon cancer stem cells, has undergone rapid development in recent years. Under physiological conditions, intestinal homeostasis is a carefully balanced and efficient interplay between stem cells, their progeny and the microenvironment. These interactions regulate the astonishingly rapid renewal of the intestinal epithelial layer, which consequently puts us at serious risk of developing cancer. Here we highlight the microenvironment-derived signals that regulate stem-cell fate and epithelial differentiation. As our understanding of normal intestinal crypt homeostasis grows, these developments may point towards new insights into the origin of cancer and the maintenance and regulation of cancer stem cells.**

**T**he astounding renewal capacity of the intestinal epithelium<sup>1</sup> has made the intestine one of the favourite tissues in which to study stem-cell regulation. The fact that almost all epithelial cells in the intestinal lining are replaced on a weekly basis puts great demands on the cellular organization of this tissue, and also puts it at serious risk of malignant conversion. Indeed, colorectal cancer (CRC) is one of the most common human cancers worldwide, with approximately 1.2 million new cases every year<sup>2</sup>. Homeostasis of the intestinal epithelium is maintained by an intestinal stem cell (ISC) compartment that resides at the bottom of the crypt, safely tucked away from the shear stresses and potentially toxic agents that pass through the intestinal tract (Box 1 Figure). These ISCs are at the top of a cellular hierarchy and are crucial for the renewal of the differentiated progeny within the intestinal layer.

The intestinal renewal system is tightly controlled and depends on the spatial organization of signals that emanate from supportive mesenchymal cells, as well as from differentiated epithelial progeny. Intriguingly, recent evidence suggests that intestinal cancers may still contain a hierarchical organization, with cancer stem cells (CSCs) at the apex<sup>3</sup>. From the seminal work of Fearon and Vogelstein it is clear that CRC develops as a stepwise accumulation of genetic hits in specific genes and pathways<sup>4</sup>. The CSC theory refines this model further and suggests that the actual tumorigenic capacity of individual cancer cells may be influenced by homeostatic signals derived from their microenvironment<sup>5</sup>. These findings are especially exciting in the light of recent developments that have increased our comprehension of the regulatory mechanisms that control ISCs, and have resulted in new tools to identify and localize ISCs (see Box 1). Although we clearly do not fully grasp the complete spectrum of signals and interactions at this point, our understanding of normal crypt homeostasis and the identification of markers that define ISCs are providing intriguing insight into the organization of intestinal cancers.

In this Review, we discuss the current ideas surrounding the identity of the ISC and the microenvironment-derived signals that regulate crypt homeostasis. In addition, we discuss the origin of cancer and the role of ISCs and CSCs, and present evidence that points to a distinctive role for the microenvironment in the onset of cancer and the maintenance of CSCs.

## Stem-cell homeostasis and morphogenetic pathways

Until relatively recently, ISCs were a rather elusive entity at the bottom of the intestinal crypt and could be identified only by indirect measurements. The discovery of ISC markers has partly changed this,

but different markers point to distinct cells within the crypt<sup>6–10</sup>. In Box 1, we detail this debate, as well as the organization of the intestinal crypt and villi. In short, the marker LGR5 points to the crypt base columnar cells located in between the Paneth cells at the crypt bottom<sup>6</sup>, whereas the markers BMI1 and TERT identify the +4 position in the crypt (Box 1 Figure), just above the Paneth cells<sup>7–9</sup>. The existence and interdependency of these different types of ISC remain a matter of debate (Boxes 1 and 2). Almost 40 years ago, the unitarian theory proposed that crypts are monoclonal populations derived from a single ISC<sup>11</sup>. More recent data, however, point to a model in which several ISCs within a crypt stochastically drift towards clonality<sup>12,13</sup>. Whether this involves different ISC types is not known (see Box 2 for a discussion on neutral drift and monoclonality). Regardless of this dispute about ISC identity, there is a consensus that ISCs reside in a niche that provides the cells with essential signals, with the morphogenetic pathways (WNT, Notch, bone morphogenetic proteins (BMPs) and Hedgehog) having a prominent role. Several excellent reviews cover these pathways<sup>14–16</sup>, so here we mainly highlight the spatial regulation instigated by the epithelial cells and their microenvironment, while an overview of the morphogenetic signalling pathways is shown in Fig. 1.

A wide range of evidence indicates that the WNT pathway (Fig. 1a) has a crucial role in intestinal proliferation and ISC maintenance. For instance, inactivating mutations in adenomatous polyposis coli (*Apc*) or activating mutations in  $\beta$ -catenin (*Ctmb1*) — both key WNT signalling factors — drive intestinal hyperplasia<sup>16</sup>. Similarly, overexpression of the WNT activator R-spondin-1 induces ISC expansion *in vivo*<sup>17</sup>. Conversely, transgenic expression of the WNT inhibitor DKK1 (ref. 18) or deletion of the transcription factor TCF4 (ref. 19) results in a block in WNT signals and subsequent attrition of the epithelial layer. Although WNT proteins are expressed in a highly complex fashion by both epithelial and mesenchymal cells<sup>20</sup>, nuclear localization of  $\beta$ -catenin is only observed at the crypt bottom<sup>21</sup>. More recent data show that Paneth cells residing next to ISCs are one of the main sources of WNT3a, and they consequently spatially constrain ISCs to the crypt bottom<sup>22</sup>. Moreover, deletion of Paneth cells decreases the number of ISCs in the crypt<sup>22</sup>. The Paneth cell signals thus dictate the size of the stem-cell pool. Whether this is a dominant trait of Paneth cells — that is, whether progenitor cells can regain ISC characteristics when placed between or next to Paneth cells — is not known, but it is clear that an active WNT signal is crucial for ISC maintenance.

<sup>1</sup>Laboratory of Experimental Oncology and Radiobiology (LEXOR), Center for Experimental Molecular Medicine, Academic Medical Center, Amsterdam 1105AZ, The Netherlands.

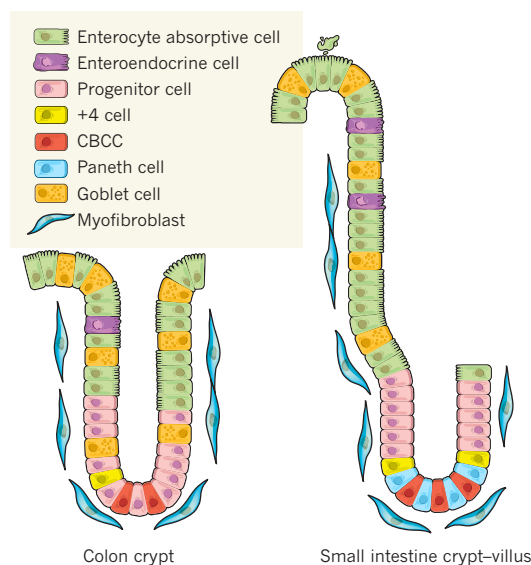
## BOX 1

# Localization and identification of intestinal stem cells

The intestinal epithelial layer consists of several differentiated cell types and is lined with mesenchymal cells<sup>14</sup> (see Box 1 Figure). The bottom of the small intestinal crypt contains Paneth cells and ISCs, whereas the remainder of the crypt is largely occupied by transit-amplifying cells, which are estimated to divide twice a day and are key to the rapid renewal of the epithelium<sup>14</sup>. At the top of the crypt, proliferation halts, and cells differentiate into either secretory (goblet, Paneth and enteroendocrine) cells or enterocytes. Although the colon lacks villi, the organization is roughly the same, except that Paneth cells are not present and differentiated cells occupy a large part of the crypt.

The unitarian theory, proposed by Cheng and Leblond in 1974, suggests that all different intestinal cell types are derived from a single stem cell<sup>11</sup>. Using a sophisticated tracing technique, they defined the crypt base columnar cells (CBCCs), named after their appearance and location between the Paneth cells, to be the ISC<sup>11</sup>. However, label-retention assays performed in the 1960s and optimized extensively since<sup>93</sup> make a convincing plea that ISCs are not located between Paneth cells and are instead located just above the Paneth cells at the +4 position<sup>93</sup>. This latter assay is based on the idea that ISCs retain a DNA label because of an immortal DNA strand that they hold onto after every division to maximize genome integrity<sup>93</sup>. The identification of +4 cells as ISCs is supported by lineage-tracing experiments using a mouse strain expressing a Cre recombinase–oestrogen receptor fusion protein (Cre<sup>ER</sup>) within the *Bmi1* locus<sup>9</sup>. This *Bmi1*-Cre<sup>ER</sup> mouse, when crossed with a *lacZ* reporter mouse, specifically activates the *lacZ* gene in the +4 position. More importantly, this irreversible genetic mark moves up and down the crypt in the days after Cre-mediated recombination to stain all differentiated cell types, suggesting that BMI1<sup>+</sup> +4 cells behave as ISCs<sup>9</sup>. This is further substantiated by the use of a transgenic mouse telomerase reverse transcriptase (*Tert*)–green fluorescent protein (GFP) construct that also marks cells at the +4 position<sup>7</sup>, as well as by lineage-tracing using *Tert*-driven Cre<sup>ER</sup> mice<sup>8</sup>. Although these data strongly argue for the +4 cell as an ISC, they are not entirely consistent, as the TERT-expressing cell has been shown to be mainly quiescent and radioresistant, whereas BMI1<sup>+</sup> and label-retaining cells divide every 24 hours and are radiosensitive. More importantly, using similar lineage-tracing techniques, another group came to completely different conclusions. On the basis of the idea that WNT is a crucial factor in ISC homeostasis, the researchers make use of specific WNT targets to mark stem cells in the crypt and find *Lgr5* to be a reliable marker<sup>6</sup>. Knock-in constructs that allow expression of GFP and Cre<sup>ER</sup> from the *Lgr5* locus show that LGR5 expression is confined to CBCCs, and that these cells give rise to the variety of epithelial cells present in crypts, proving that CBCCs function as ISCs as well<sup>6</sup>. Further studies show several other specific markers for these cells, such as *Olfm4* and *Ascl2* (ref. 94). The transcription

factor ASCL2 is of particular importance because its deletion results in the complete loss of LGR5<sup>+</sup> ISCs, whereas transgenic *Ascl2* expression induces crypt hyperplasia<sup>94</sup>. Notably, BMI1 does not seem to have a significant role in LGR5<sup>+</sup> ISCs as *Bmi1*-knockout mice show normal CBCC morphology, maintain ASCL2 and OLFM4 expression and, above all, have a normal intestinal epithelium<sup>94</sup>. It therefore remains to be determined whether and how BMI1<sup>+</sup> +4 ISCs and LGR5<sup>+</sup> ISCs relate to each other. Interestingly, recent data indicate that TERT-expressing ISCs are quiescent, reside at the +4 location and can generate LGR5<sup>+</sup> ISCs<sup>8</sup>. Although this suggests that these different ISC types may act in a hierarchical fashion (see also Box 2), the observed generation of LGR5<sup>+</sup> cells from TERT<sup>+</sup> cells is not very efficient, and the model clearly contrasts with observations that claim that the LGR5<sup>+</sup> ISCs show the highest telomerase activity in the crypt<sup>95</sup>.



**Box 1 Figure | The organization of the colon crypt and the small intestinal crypt–villus.** Both the colon crypt (left) and the small intestinal crypt (right) contain a stem-cell compartment at the crypt bottom. CBCCs and the +4 stem cell have been indicated to be present between and just above the Paneth cells, respectively. Of note, Paneth cells are not detected in the colon, yet a Paneth-like cell has been suggested to be present at the crypt bottom. All four lineages (three in the colon) — enterocytes, Paneth cells, goblet cells and enteroendocrine cells — appear in different, but set ratios. Paneth cells move down to the bottom and are long-lived, whereas other lineages move up and are shed (a few days later) into the lumen while undergoing apoptosis. Rare cell types reported to exist in crypts, such as tuft cells, are not shown.

Activation of the Notch pathway (Fig. 1b) normally depends on direct cell-to-cell contact, but secreted ligands exist. Notch interacts with ligands from the delta and jagged family and shows classical feedback inhibition<sup>14</sup>. Cells that receive signals through Notch downregulate their own Notch ligands and thus deprive their neighbouring cells of Notch-activating signals. This mechanism allows easy regulation of different cell-fate decisions for neighbouring cells. In the intestine, Notch activity determines lineage decisions between enterocyte and secretory cell differentiation. Inhibition of the Notch pathway results in a massive increase in goblet cells, whereas its activation results in goblet-cell depletion<sup>23,24</sup>. However, Notch seems to have dual functions in the crypt, as it acts together with WNT to maintain the proliferative speed and deletion or activation of the pathway, which significantly

affects crypt homeostasis<sup>23–25</sup>. Similar to other developmental systems, data suggest that the ISC microenvironment delivers Notch-activating signals to maintain 'stemness', which is consistent with the observation that Paneth cells express Notch ligands<sup>22</sup>. However, it is important to note that it is unclear whether ISCs require Notch to retain stemness or whether Notch instead acts on transit-amplifying cells. Nevertheless, in general, the data strongly support a model in which Notch directs proliferation when WNT signal activity is high, and directs enterocyte differentiation when WNT activity levels drop towards the top of the crypt.

BMP belongs to a more complex family of ligands comprising BMP and transforming growth factor- $\beta$  (TGF- $\beta$ ) family members, which share intracellular signalling through the SMAD proteins<sup>15</sup> (Fig. 1c).



## BOX 2

## Crypt monoclonality and stem-cell neutral drift

Intestinal crypts are monoclonal in origin, which concurs with the idea that they are organized in a strict hierarchical fashion. Monoclonality is achieved within 2 weeks of birth in a process called purification<sup>96</sup>. When a mutation is introduced, crypts go through a mosaic phase, but quickly become monoclonal again (reviewed in ref. 93). These observations suggest that a mutation in a particular cell can quickly spread throughout the crypt, which has implications for cancer. Moreover, the data suggest that a single ISC generates the complete crypt, giving rise to a monoclonal lineage. However, this model has been modified to introduce the concept of neutral drift<sup>8,97</sup>, a term derived from population genetics. In a stem-cell context, this describes how several functionally identical ISCs generate a monoclonal crypt by assuming that at every division ISCs stochastically generate zero, one or two daughter ISCs. When zero daughter ISCs are formed, this specific clone is lost and quickly replaced by a neighbouring ISC, hence explaining how several ISCs can, by neutral drift alone, generate monoclonal crypts. Sophisticated genetic-tracing strategies using either *Lgr5*-driven 'confetti' mice or *Ah-Cre<sup>ER</sup>* mice further verified this concept, leading to the conclusion that several ISCs within a crypt show a non-hierarchical organization that is subject to neutral drift<sup>12,13</sup>. Importantly, the onset of cancer through *APC* mutations greatly influences crypt dynamics, because this tumorigenic stem-like cell is suggested to disregard neutral drift and possess the capacity to quickly repopulate the crypt with its own progeny, generating early aberrant crypt foci<sup>98</sup> (Fig. 2).

Although the above-described model satisfactorily explains the mechanism by which monoclonality is attained, a few puzzling issues remain. The first is the observation that crypts are polyclonal at birth, before purification<sup>96</sup>. Apparently, the neutral drift is not occurring prenatally. The second issue is the purification speed of the adult colon, which is faster than in the small intestinal crypts. Yet colon ISC proliferation, a crucial determinant for the pace of neutral drift, is reportedly slower<sup>99</sup>. In the current, stochastic drift, model this can be explained only if the number of stem cells is significantly lower in the colon. An alternative explanation for these observations could be that a slow cycling or quiescent ISC sits on top of the hierarchical organization in adult crypts, similar to the organization of the haematopoietic stem-cell compartment<sup>100</sup>. This ISC would be activated under extreme stress conditions such as irradiation. In agreement with such a model is the observation that TERT-expressing quiescent ISCs can give rise to LGR5<sup>+</sup> ISCs after radiation damage<sup>8</sup>. Importantly, this also sheds new light on crypt monoclonality, which could thus be a combination of neutral drift of LGR5<sup>+</sup> ISCs and repopulation by rare 'master' ISCs. It may prove insightful if lineage-tracing experiments with *Lgr5-Cre<sup>ER</sup>* mice are followed by sublethal irradiation to kill the LGR5<sup>+</sup> ISCs. ISC and crypt regeneration through fission or by a previously quiescent ISC will show distinct dynamics and probably provide insight into the relationship between CBCCs, +4 ISCs and quiescent ISCs.

BMP2 and BMP4 are expressed by mesenchymal cells in the intestine and are suggested to halt proliferation at the crypt-villus border, allowing differentiation. In agreement, *Bmpr1a*-deficient mice (which lack a BMP receptor)<sup>26</sup> or mice overexpressing the BMP inhibitor noggin<sup>27</sup> show hyperproliferation, as well as crypt fission. To coordinate the segregation between proliferation and differentiation, BMPs are active at the top of the crypt, where differentiation occurs. BMPs are also produced at the crypt bottom, but here they are kept in check by BMP inhibitors (such as noggin) that are specifically expressed by the mesenchyme in the ISC region. Notably, BMPs also have dual functions and are involved in lineage fate decisions towards secretory cells<sup>28</sup>. Through BMPs and noggin, the mesenchymal microenvironment thus secures a spatial organization in the crypt (Fig. 1c).

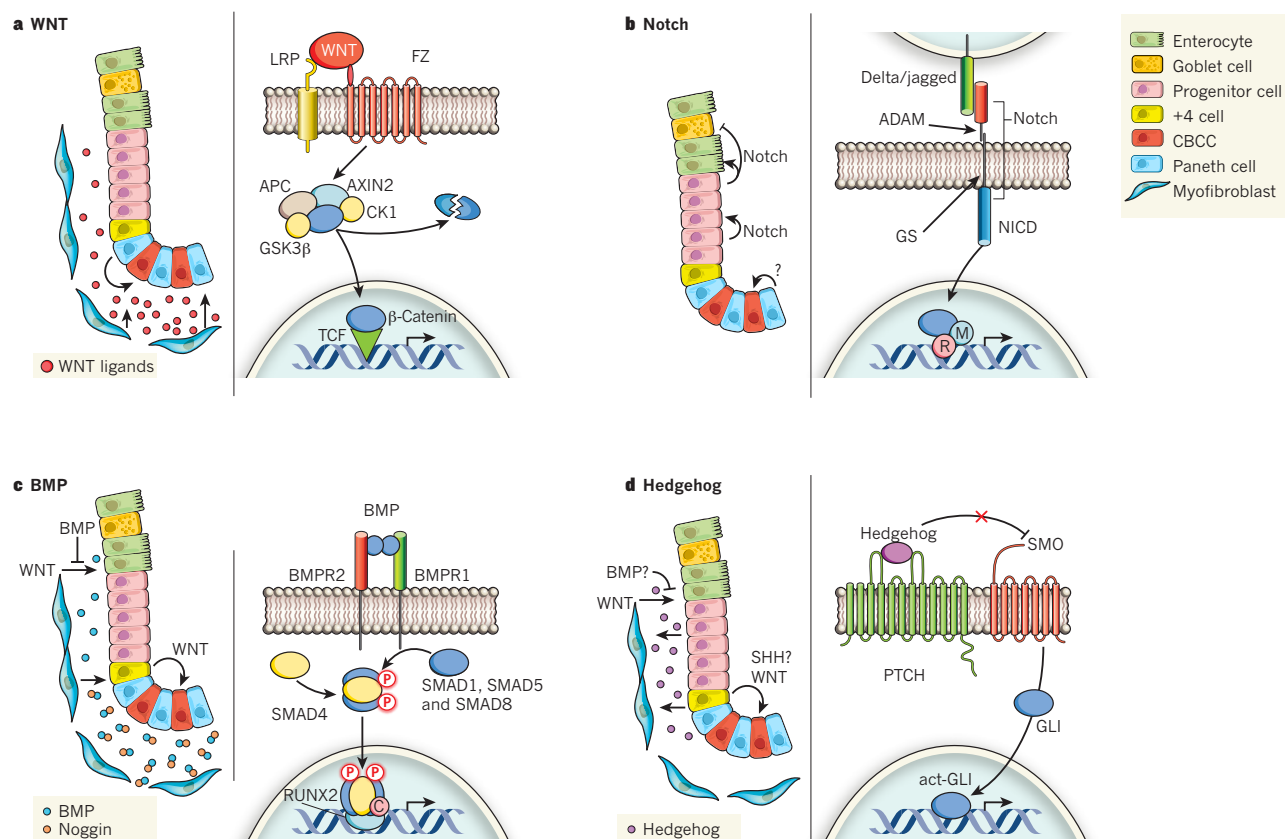
The role of the Hedgehog pathway (Fig. 1d), which acts through the membrane proteins smoothened and patched<sup>29</sup>, in intestinal homeostasis is more confusing. Conditional deletion of patched, a negative regulator of the pathway, inhibits WNT in the small intestine and leads to premature enterocyte differentiation<sup>30</sup>. Physiologically, it seems that Indian hedgehog (IHH), which is mainly expressed by differentiated epithelial cells, signals to the mesenchyme, where it is thought to induce BMP secretion<sup>31,32</sup>. In contrast to IHH, the role of Sonic hedgehog (SHH) is less clearly defined. SHH is reported to be expressed in a single cell in the crypt at the +4 position, which demonstrates ISC features<sup>33</sup>, but further studies are needed to define this cell and its function better.

Together, the current data paint a picture in which the micro-environment of ISCs, composed of direct progeny, mesenchymal cells and probably extracellular matrix components, organizes into a complex interaction of morphogenetic signals that each has a crucial role in crypt maintenance (Fig. 1 and Box 1). This complexity explains why *in vitro* recapitulation of the crypt architecture has proven extremely difficult to achieve. However, the insight into crypt homeostasis has allowed two independent groups to devise culture methods that successfully generate crypt structures, which can be sustained for long periods of time *in vitro*<sup>34,35</sup>. Both culture systems require a solid matrix (collagen

or Matrigel) and are strongly potentiated by R-spondin-1, which enhances WNT signalling. In both systems, cells seem to organize into crypt-like structures containing ISCs and transit-amplifying cells, as well as differentiated cells. Interestingly, a completely different niche-dependency was observed in the two culture systems. Whereas the collagen cultures depend on mesenchymal myofibroblasts<sup>34</sup>, these are redundant in Matrigel<sup>35</sup>. It is important to note that the latter cultures require noggin<sup>35</sup>, probably preventing BMP-driven differentiation of ISCs and transit-amplifying cells. Whether noggin or, alternatively, growth factors present in Matrigel replace the myofibroblasts remains to be established, but it is clear that these cultures have enormous potential for further study. For instance, as both BMI1<sup>+</sup> and LGR5<sup>+</sup> cells are present in the crypt cultures<sup>34,35</sup>, this should allow *in vitro* lineage-tracing and thereby provide an avenue to address the relationship between different ISCs in easily accessible systems. At this point, the Matrigel culture system has been used to show that even a single LGR5<sup>+</sup> ISC can expand and self-organize into a crypt-like structure<sup>22</sup>. This process is greatly enhanced in the presence of WNT-ligand-producing Paneth cells<sup>22</sup>, validating the idea that Paneth cells support ISCs, and simultaneously confirming that ISCs require input from a niche. Undoubtedly, these culture methods will in the near future provide us with more direct insight into the role of mesenchymal cells and the distinct morphogenetic pathways in ISC maintenance and differentiation.

### Intestinal cancer

Despite stringent homeostatic maintenance in the intestine, the high number of patients with CRC indicates that these regulatory mechanisms often fall short in protecting against malignant transformation. Both environmental and genetic risk factors have been defined for CRC, and, not surprisingly, deregulation of morphogenetic pathways plays a key part in cancer development. Environmental factors include a Western diet and a history of inflammatory bowel disease<sup>36</sup>, whereas a genetic component has been clearly defined by genome-wide association studies<sup>37,38</sup>. In addition, prominent genetic



**Figure 1 | Crypt homeostasis.** **a**, WNT ligands around the base of the intestinal crypt signal through a complex of frizzled (FZ) and low-density lipoprotein receptor-related protein (LRP) receptors. Canonical WNT regulates  $\beta$ -catenin localization through a destruction complex that contains APC, AXIN2, glycogen synthase kinase-3 $\beta$  (GSK3 $\beta$ ) and casein kinase 1 (CK1), and directs the phosphorylation and subsequent degradation of  $\beta$ -catenin. In the presence of WNT ligands, this complex disassembles, allowing the accumulation and translocation of  $\beta$ -catenin to the nucleus, where it drives the transcription of WNT target genes aided by TCF and lymphocyte enhancer factor-1 transcription factors. **b**, Notch cooperates with WNT to drive proliferation, and is involved in lineage-fate decisions. This straightforward pathway is active in cell-to-cell contact. Delta and jagged ligands on the surface of one cell activate Notch receptors on a neighbouring cell. This is followed by two proteolytic events catalysed by a disintegrin and metalloprotease (ADAM) and  $\gamma$ -secretase (GS) proteases, which release Notch intracellular domain (NICD). NICD translocates to the nucleus, where it drives the transcription of Notch gene targets aided

by mediators such as recombining binding protein suppressor of hairless (R) and mastermind-like protein 1 (M). **c**, BMP proteins, produced mainly by stromal cells, counteract proliferative WNT signals and thereby halt proliferation and drive differentiation. At the crypt bottom, BMPs are blocked by noggin, which specifically binds BMPs and prevents receptor interaction. Signalling by BMPs depends on the heterodimerization of the BMPR1 and BMPR2 receptors, leading to phosphorylation of SMAD1, SMAD5 and SMAD8. Phosphorylated (P) SMADs associate with SMAD4, translocate to the nucleus and drive the transcription of BMP target genes aided by RUNX2 and a cofactor (C). **d**, Hedgehog proteins relay signals between epithelial cells and the mesenchyme. Hedgehog has been shown to counteract WNT-driven epithelial proliferation, potentially through BMPs. The role of SHH in ISC biology remains unclear. Hedgehog signalling relies on the interaction between patched (PTCH) and smoothened (SMO). Patched represses smoothened, but is blocked when bound to Hedgehog. Derepressed smoothened activates GLI transcription factors (act-GLI), which translocate directly to the nucleus and drive the transcription of Hedgehog target genes.

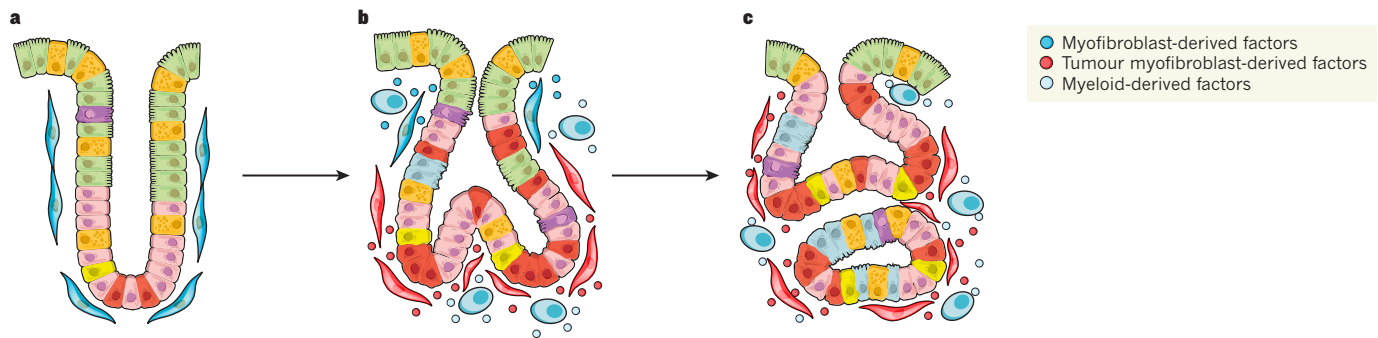
predispositions are found in several familial CRC syndromes, such as familial adenomatous polyposis (FAP) and hereditary non-polyposis CRC (HNPCC or Lynch syndrome)<sup>37,38</sup>. Patients with FAP develop hundreds of colonic polyps early on in life, and their lifetime risk of developing CRC is almost 100%<sup>39</sup>. The genetic defect that underlies this syndrome is a heterozygous mutation in the APC gene — a crucial negative regulator of the WNT pathway<sup>40</sup> (Fig. 1a). Notably, the vast majority of sporadic CRC cases also carry WNT pathway mutations (85% APC and 10%  $\beta$ -catenin), highlighting the importance of this pathway in CRC. On the basis of work by Fearon and Vogelstein, the prevailing model is that CRC develops owing to an accumulation of mutations, each defining a different step in the adenoma–carcinoma sequence<sup>4</sup>. In this sequence, the first hit that induces the transition from normal to polypoid tissue is seen in the WNT pathway, whereas progression to adenomas and carcinomas depends on, for example, activating mutations in the RAS pathway and inactivation of p53, SMAD4 and PTEN<sup>39</sup> (Fig. 2). Although this mutation-driven model is supported by a range of experimental data, several nuances need to be made to explain the current observations

fully. First, tumorigenicity induced by mutations is proposed to be different when these are introduced in ISCs compared with transit-amplifying or differentiated cells, which has led to the idea that ISCs are the cell of origin in cancer. Second, microenvironmental influences on CRC need to be placed within this scheme of events, as it is now well established that environmental factors are an enabling characteristic promoting tumour initiation and growth<sup>41</sup>, and this is especially evident in CRC. Third, the existence of cell-to-cell variation in the grade of differentiation within CRC as proposed by the CSC hypothesis<sup>3</sup> needs to be evaluated and placed within the context of the mutation-driven model.

### Stem cells and the origin of cancer

As mentioned earlier, the sequence of events in CRC has been intensively studied using a variety of mouse models. The most frequently used model is the *Apc*<sup>Min</sup> mouse, which was generated by a random mutagenesis screen<sup>42</sup>. Similar to patients with FAP, this mouse strain contains a heterozygous truncating mutation in the APC gene, and develops dozens of polyps and small adenomas





**Figure 2 | Stem cells and the environment in the adenoma–carcinoma sequence.** **a**, Normal organization of the intestinal crypt. **b**, After the loss of wild-type *APC* or  $\beta$ -catenin mutation, the transformation of healthy crypts towards an adenoma starts. This is accompanied by several changes in crypt appearance and behaviour. First, cells show a more immature phenotype and a higher proliferative index (more red and pink cells with irregular shape). Second, crypt fission, a physiological process in which a crypt splits, is observed and results in expansion of the pre-malignant clone. Third, the attraction of myeloid cells (round light blue), producing factors such as interleukin-6 and tumour-necrosis factor- $\alpha$  (light blue dots) is detected

and promotes carcinogenesis. Although normal myofibroblasts (blue) are still present, factors produced by premalignant and infiltrating cells activate myofibroblasts (orange). The myofibroblasts produce, among other factors, large amounts of hepatocyte growth factor (HGF), which promotes dedifferentiation (orange dots). **c**, The accumulation of other genetic lesions, in *RAS* and *PTEN*, for example, induces progression towards an invasive growing CRC. At this stage, stromal cells become an even more pronounced part of the tumour through the production of factors that further promote tumour progression (orange dots). Note the continued presence of several types of differentiated (tumour) cell throughout the whole sequence.

in the intestine. Although this strain has been available for a long time, the cell of origin for cancer formation in this model of CRC has remained obscure. Both a bottom-up theory, in which ISCs are the cell of origin, and a top-down theory, in which a progenitor or differentiated cell is the first transformed cell, have been suggested. Evidence for the top-down model of CRC development relies almost completely on histopathological observations. By contrast, the bottom-up theory recently received strong genetic support, as ISC-specific deletion of both functional *Apc* alleles using *Bmi1*-, *CD133* (also known as *Prom1*)- and *Lgr5*-Cre recombinase mice leads to the very rapid development of full adenomas<sup>9,10,43</sup>. Notably, in a parallel approach, deletion of functional *Apc* in short-lived progenitor or differentiated cells resulted in only sporadic and slow-developing adenomas<sup>43</sup>. Similar conclusions on the role of normal stem cells as the cell of origin of tumours have been generated using mouse models for chronic myeloid leukaemia<sup>44</sup>, prostate carcinoma<sup>45</sup> and glioblastoma<sup>46</sup> (see ref. 47 for review). However, several important issues on the role of microenvironmental factors, the speed of tumour development and the true nature of human tumours need to be considered before translating this knowledge to human CRC.

### The adenoma–carcinoma sequence and polyclonality

Although the models for ISC-driven tumorigenesis provide valuable insights, the speed at which adenomas form in these systems is much faster than in humans. The speed of tumour formation can be attributed to a simultaneous deletion of both *Apc* alleles, but this rapid generation of adenomas is unlikely to be a true mirror image of human disease. For example, in large human case studies, very early adenoma precursor lesions or aberrant crypt foci (ACF), which mostly occur through *APC* mutations, are much more abundant than adenomas<sup>48</sup>. This indicates that the transition from an *APC* mutant ACF stage to a full adenoma is a slow process, or that most ACF never progress to an adenoma. In agreement, studies on a unique mosaic patient with FAP show that most of this patient's adenomas are polyclonal in nature, whereas the early lesions (monocryptal ACF) are all monoclonal<sup>49</sup>. This suggests that the transition from ACF to adenomas may be a much more complex process than simple expansion. In this light, it is interesting to note that the deletion of *Apc* in mouse intestinal progenitor cells gives rise to micro-adenomas that remain present until later age<sup>43</sup>, in contrast to the idea that such clones would be pushed to the top of the villi and discarded into the gut lumen. In some cases, those lesions even progress to adenomas<sup>43</sup>. Arguably, this sequence of events provides a more realistic model for human disease, at least from a timing perspective.

### APC mutations are carefully selected

Despite the fact that the deletion of both *Apc* alleles results in rapid adenoma formation, WNT activity levels in human CRC seem to be strongly regulated during both initiation and progression. This is partly due to the genetic make-up of developing lesions and is partly induced by the microenvironment. For instance, loss of the normal *APC* allele in individuals with FAP is a non-random event, because the nature of germline *APC* mutation has been shown to influence the type of the second *APC* hit<sup>50,51</sup>. In essence, it seems that the shorter the germline-truncated *APC* protein is, the longer the somatic-mutated *APC* protein will be, and vice versa. As the length of truncated *APC* is directly linked to the ability to prime  $\beta$ -catenin for degradation<sup>52</sup>, this led to the 'just right' signalling model, in which only a specific range of WNT signalling activity levels (not too high, not too low) is associated with transformation of intestinal epithelial cells<sup>50</sup>. This observation suggests a careful balance in WNT activity during CRC initiation, and it is therefore not clear whether simultaneous deletion of both mouse alleles in ISCs is representative of human disease.

### Adenoma formation is influenced by the microenvironment

The above-described ISC-driven models suggest an almost cell-autonomous induction of adenomas. However, in the *Apc*<sup>Min</sup> mouse model, regulation by environmental factors is clearly observed. As an example, the administration of dextran sulphate sodium (DSS), which induces intestinal inflammation, results in a strong increase in polyp formation in *Apc*<sup>Min</sup> mice, indicating that microenvironmental factors have an important role<sup>53</sup>. Whether DSS-induced inflammation can also support adenoma formation when *Apc* is deleted in progenitor cells is unclear, but if this were the case it would suggest that the cell of origin in CRC depends on the model chosen.

Although this cell-of-origin discussion seems at first sight to be an academic issue that will not directly impinge on patients with cancer, there is good reason to understand the history of CRCs. For instance, it may be crucial to understand whether different types of CRC (microsatellite instability, mucinous or neuroendocrine) develop owing to different genetic hits or from a distinct cell of origin. More importantly, with respect to prevention, knowing the cell of origin and understanding the signals that either maintain lesions as small ACF or allow them to progress to full adenomas and subsequently carcinomas is of vital importance. We can conclude that *APC* deletion in ISCs leads to the very rapid onset of adenomas<sup>43</sup>, indicating that ISCs are more prone to full transformation. This is probably due to their capacity

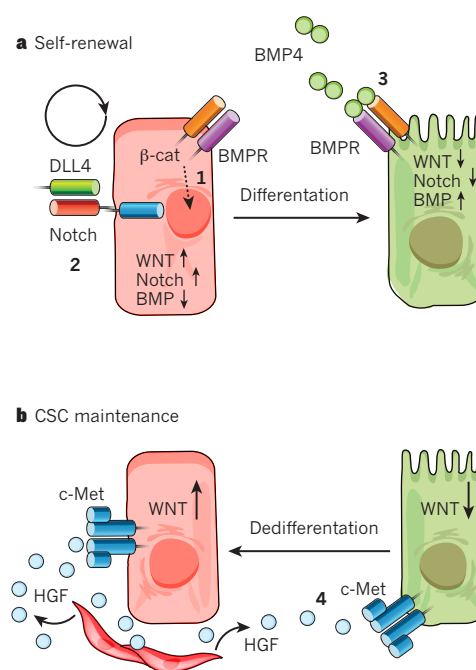
to self-renew, but caution is needed to translate these findings to spontaneous adenoma development, in which the microenvironment and other genetic hits are crucial determinants.

### Genetic and microenvironmental influences in CRC

To gain more insight into the development of CRC and the genetic and environmental modulations of this process, several mouse models have been generated that extend the *Apc*<sup>Min</sup> model of CRC. In most of these models, tumorigenesis occurs in the small intestine, whereas in humans the disease appears exclusively in the colon. The reason for this discrepancy is not well defined, but it is notable that humans rarely develop small intestinal malignancies, although homeostatic signals and renewal speed are largely identical between colon and small intestine. Besides a difference in location, *Apc*<sup>Min</sup> mouse adenomas also rarely progress to full carcinomas<sup>54,55</sup>, which may be a timing issue because progression in humans is estimated to take approximately 2–10 years — much longer than the lifespan of a mouse. To model human CRC better, mouse strains containing different types of truncated *Apc* gene, representing the variety of human *APC* mutations, have been generated, and these show subtle differences in the number and location of adenomas, indicating that the dosage of WNT signalling is crucial<sup>55</sup>. Notably, *Apc*<sup>+/-</sup> mice with an extra mutation in the homeobox gene *Cdx2*, a regulator of colon epithelial cell differentiation, shift most polyps to the colon<sup>56</sup>. However, progression to invading carcinomas in such mice is not observed. Therefore, the next steps in the adenoma–carcinoma sequence are modelled by introducing further mutations in addition to *Apc*. For example, *Apc*<sup>+/-</sup> mice that also contain an activating *Ras*<sup>V12G</sup> mutation show more dysplasia and enhanced invasive growth<sup>57,58</sup>. Similar results have been obtained by deleting *Pten* on an *Apc*<sup>+/-</sup> background<sup>59</sup>. Together, these data support the notion that sequential accumulation of genetic mutations underlies CRC development, as proposed previously<sup>4</sup>. However, these studies also show that the microenvironment is a determining factor in colorectal tumorigenesis. The initial search for genetic modifiers of the *Apc*<sup>Min</sup> phenotype indicated that secreted phospholipase A<sub>2</sub>, an enzyme involved in inflammatory responses, potentiates hyperplasia<sup>60,61</sup>. Similarly, microenvironmental tumour control was deduced from the *cis-Apc/Smad4* mouse<sup>62</sup>. The tumours that arise in this mouse strain carry an *Apc* mutation in combination with a lack of *Smad4*, which renders the mice unresponsive to differentiation-inducing BMP signals from the microenvironment. These mice have highly invasive adenocarcinomas, characterized by extensive stromal proliferation<sup>62</sup>. More recent findings indicate that this invasive phenotype depends on the recruitment of immature myeloid cells to the microenvironment<sup>63</sup>. Interestingly, CRC progression in humans is also associated with a loss of BMP signalling through inactivation of either SMAD4 or BMPR2 (ref. 64). The loss of microenvironment-derived BMP-mediated control of epithelial proliferation thus seems to promote progression.

More direct evidence for microenvironmental control of CRC comes from mouse models for human juvenile polyposis and Peutz–Jeghers syndrome, which are both characterized by hamartomatous polyps. Blockade of microenvironmental BMP4 signals by ectopic expression of noggin (Fig. 1c), but also *Smad4* deletion in mouse T cells, results in juvenile-polyposis-like polyp formation<sup>27,65</sup>. Similarly, a germline *Lkb1* (also known as *Stk11*) mutation or an *Lkb1* mutation specifically in mesenchymal cells induces a Peutz–Jeghers-like syndrome in mice, with a prominent stromal compartment<sup>66</sup>. This points to a dominant role for the microenvironment in this CRC subtype.

In humans, the strongest support for environmental control of CRC comes from the long-standing observation that chronic inflammation in patients with Crohn's disease or ulcerative colitis predisposes them to cancer initiation in the gut<sup>67</sup>. In CRCs that are not associated with inflammation, there is also clear histopathological evidence for the function of immune cells in the stroma of malignancies. In addition, the administration of non-steroidal anti-inflammatory drugs lowers the risk of CRC-specific mortality in humans<sup>68</sup>, and specific cyclooxygenase 2



**Figure 3 | Regulatory signals of colon CSCs provide new therapeutic targets.** **a**, Colon CSCs receive a multitude of environmental cues. Signals that help to maintain CSCs include Notch and WNT. DLL4 stimulates Notch receptors on neighbouring cells and, together with β-catenin (β-cat), directs an immature transcription profile that promotes self-renewal. BMP4 counteracts this self-renewal activity by binding to BMP receptors on CSCs, thereby interfering with WNT signalling and subsequently promoting differentiation. **b**, Colon CSCs *in vivo* are found intimately associated with HGF-producing myofibroblasts. HGF maintains colon CSCs in a stem-cell state and prevents differentiation. In addition, HGF produced by myofibroblasts can upregulate the WNT cascade in more differentiated tumour cells, thereby reinstalling CSC features (dedifferentiation). These signals that govern CSC behaviour have great therapeutic potential. Several possibilities to interfere with the self-renewal capacities of CSCs exist. These include inhibitors of the WNT pathway that prevent β-catenin-dependent transcription (1); Notch inhibitors, preventing either the ligand from interacting with the receptor or the activation of the receptor (γ-secretase inhibitors), which are currently under evaluation (2); BMPR agonists that activate differentiation programs and could be used to target CSCs (3); and inhibitors of the receptor kinase c-Met that could modulate the interaction between stromal cells and CSCs to prevent dedifferentiation (4).

inhibitors cause reductions in intestinal polyps in *Apc*<sup>Min</sup> mice<sup>69</sup>. This effect is at least partially mediated by a decrease in prostaglandin E<sub>2</sub> (PGE<sub>2</sub>) levels, which leads to a reduction in myofibroblast secretion of tumour-supporting factors such as hepatocyte growth factor (HGF) and amphiregulin<sup>70</sup>. In addition, lower PGE<sub>2</sub> levels decrease the influx of inflammatory cell types, such as tumour-associated macrophages and mast cells, which form an important tumour-promoting component of the cellular stroma<sup>71</sup>.

In mice, chronic inflammation-associated CRC is often studied using the DSS and azoxymethane (AOM) model<sup>55</sup>. Application of the carcinogen AOM alone rarely induces intestinal neoplasia in mice but, when followed by the colitis inducer DSS, frequent tumours occur in the colon. In these mice, the invading myeloid cells produce a plethora of growth factors and cytokines, of which tumour-necrosis factor-α, interleukin-6 and TGF-β seem to be crucial for tumour initiation and progression<sup>72</sup>. These pro-inflammatory factors stimulate the inhibitor of κB kinase (IKK)–nuclear factor κB (NF-κB) pathway in epithelial cells, as shown by specific IKK-β deletion, which leads to increased proliferation, reduced apoptosis and enhanced tumour incidence<sup>73</sup> and confirms a role for microenvironmental factors in the onset of disease. Similarly, MyD88 deletion decreases adenoma formation, indicating



that stimulation of Toll-like receptors by the microflora in the gut can directly enhance tumorigenesis<sup>74</sup>. These observations from both mouse models and human disease indicate that the (inflammatory) microenvironment has a prominent role in the onset and progression of CRC by supporting genetic mutations within epithelial cells (Fig. 2).

### Cancer stem cells

An important downside of modelling cancer in mouse models is the fact that the tumours that arise are relatively genetically homogeneous, whereas human CRCs are thought to contain as many as 80 mutations, which probably vary from cell to cell<sup>75</sup>. This genetic heterogeneity within a tumour generates an added dynamic interaction<sup>76</sup> that is hard to model in mice. Furthermore, the past decade has seen a shift in the way tumours are perceived, and the now widely accepted model is that tumours contain a small population of self-renewing CSCs, as well as a large compartment of more differentiated tumour cells<sup>3</sup>. It is important to realize that this hierarchy is not identical to genetic heterogeneity, but instead adds a further layer of complexity. CSCs are defined operationally by their capacity to (xeno)transplant a tumour and thereby generate a phenocopy of the original human malignancy, including all differentiated progeny<sup>77</sup>. In contrast to the more differentiated tumour cells, CSCs are thus suggested to embody the driving force within a tumour — that is, to contain the capacity to self-renew, expand and differentiate<sup>3</sup>. An increasing number of both solid and lymphoid malignancies are shown to contain CSCs (see ref. 3 for review). However, there is still considerable disagreement on the existence or importance of CSCs in human tumours. This is partly due to the lack of unique markers for these cells that can be used to directly identify CSCs in human malignancies and thereby remove the necessity to perform (xeno)transplantation assays. Moreover, our current observations indicate that CSC characteristics can also be bestowed on differentiated tumour cells when exposed to the right microenvironment<sup>5</sup> (see later). The CSC theory may thus be more complex than originally anticipated, but it is nonetheless clear that the variation in tumour cell differentiation holds important implications for our understanding of tumours. For instance, several studies have indicated that CSCs in CRC are more resistant to therapy than differentiated tumour cells are<sup>78–80</sup>, which has led to the assumption that CSCs are crucial targets in therapy and, more importantly, may be the source of relapsing tumours<sup>5,78</sup>. In addition, we believe that the current data suggest that cellular hierarchy within CRC is maintained, at least in part, by microenvironmental factors regulating stemness and differentiation, much like normal homeostasis. In agreement, mouse adenomas that originate after inactivation of APC in LGR5<sup>+</sup> ISCs do not simply show an expansion of LGR5<sup>+</sup> cells, but instead show a distinct subpopulation of tumour cells positive for this marker<sup>43</sup>. This observation suggests that a hierarchy is established in adenomas and that differentiation towards LGR5<sup>+</sup> cells occurs, potentially as a result of normal differentiation cues. This concurs with earlier observations that adenoma cells differentiate into goblet cells after Notch inhibition with  $\gamma$ -secretase inhibitors<sup>24</sup>, in line with the physiological role for Notch in proliferation and differentiation. Together, these findings suggest that mouse adenoma cells respond, at least partially, to their normal environmental cues, thereby maintaining a hierarchy within the developing tumour.

In human CRC, CSCs are defined using CD133, CD166, CD44 and CD24 cell-surface markers<sup>5,78,81–85</sup>, whereas differentiated tumour cells express markers normally present in differentiated colon epithelium. Importantly, even a single colon CSC can generate a fully differentiated tumour after xenotransplantation, proving that CSCs have a multilineage differentiation capacity that gives rise to all differentiated progeny<sup>84</sup>, but above all they seem to be hardwired (epi)genetically such that the offspring reshapes the original human tumour. This is not unique to CRCs but widely observed for different solid malignancies, and has also been shown at the single-cell level in glioblastoma<sup>86</sup>, breast cancer<sup>87</sup> and some haematopoietic malignancies. We believe that this is partly regulated by stromal cells that are rapidly attracted

and form an essential part of CRCs in mouse models, as well as human malignancies. A recent study showed that tumour cells located next to or within myofibroblast-rich regions have a much higher incidence of nuclear-localized  $\beta$ -catenin<sup>88</sup>, directly arguing for microenvironment-modulated WNT signalling. In agreement, human colorectal CSCs can be defined on the basis of high WNT signalling activity, and preferentially localize to myofibroblasts in xenotransplants<sup>5</sup>. Our data point to HGF as the myofibroblast-derived signal that, at least in part, orchestrates this intimate relationship and enhances WNT activity<sup>5</sup>. Conversely, whether myofibroblasts also receive regulatory input from epithelial tumour cells, similar to the situation in normal crypt homeostasis, remains to be determined.

These data clearly ascertain a role for the WNT pathway in CRC stemness, but evidence for the involvement of other morphogenetic pathways also exists (summarized in Fig. 3). Notch inhibition with an antibody against the Notch ligand DLL4 results in human colon CSC differentiation, reduction of CRC growth in a xenotransplantation model and chemosensitization<sup>89</sup>. Of note, this is at least in part a tumour-intrinsic DLL4 signal as it is blocked by a human-specific anti-DLL4 antibody<sup>89</sup>. Recent findings also indicate that BMP4 can aid in human CSC differentiation, much like it does in normal epithelial progenitors, and thereby induce the chemosensitivity of tumours<sup>90</sup>. Unlike normal homeostasis, in which BMP4 is expressed by the mesenchyme, BMP4 is expressed by the differentiated tumour cells and, as such, imposes a feedback mechanism<sup>90</sup>. Finally, a role for Hedgehog in CSCs and tumorigenesis has also been proposed, as progression towards metastasis seems to be associated with a switch from WNT to Hedgehog signal dependency<sup>91</sup>. However, in mouse models, both activation and inhibition of Hedgehog signalling has been shown to prevent tumour growth<sup>30,33</sup>, making the role for Hedgehog less clear-cut.

From these findings, we propose a model encompassing a hierarchical organization of CRCs, which is notably similar to normal homeostasis (Fig. 2). In this model, the Notch pathway, together with high WNT levels that result from mutation and myofibroblast-derived factors, promote cancer stemness. The BMP pathway counteracts this self-renewal mechanism and drives tumour cell differentiation (Fig. 3). Whether this concept depends on the nature of the mutations present in a specific CRC or is influenced by the stage of disease remains to be established. Nevertheless, the current data lend strong support to the influence of the microenvironment on CSC maintenance and tumour growth. In analogy to the normal ISC niche, we conclude that there is a CSC niche that is probably composed of a combination of stromal cells and more differentiated progeny, and delivers crucial signals to the CSCs.

### Outlook

Understanding the signals that regulate tumour maintenance obviously serves one major purpose — the improvement of cancer therapies (for examples, see Fig. 3). Because colon CSCs seem subject to a careful interplay of extracellular cues, in the coming years this idea should be evaluated in defined tumour models that not only depend on genetically modified mice, but also extend findings to genetically heterogeneous human tumours containing a hierarchical organization. Only then can we accurately assess whether morphogenetic pathways and factors derived from stromal myofibroblasts modulate cancer stemness and subsequently influence tumour growth (Fig. 3). In addition, such models should provide us with more insight into the flexibility of the hierarchical organization in tumours. The current model for normal intestinal homeostasis is that once epithelial cells have lost their stemness, they will terminally differentiate. However, it is clear that dedifferentiation can easily occur in tumour settings. That is, under the influence of stromal myofibroblasts, more differentiated tumour cells can reacquire CSC features<sup>5</sup>, suggesting that the CSC phenotype is more fluid than initially proposed. Similar observations have been derived for breast cancer cells, which, after induction of the epithelial–mesenchymal transition, generate CSCs from more differentiated

tumour cells<sup>92</sup>. Such flexibility challenges the idea that targeting CSCs would be sufficient for efficient tumour control. The dominant role of microenvironment-derived HGF in the generation of CSCs in CRC indicates that exogenous factors derived from the mesenchyme can provide potent tumorigenic signals<sup>5</sup>. As CSCs are more resistant to therapy, it is not difficult to appreciate how the induction of CSCs by the microenvironment directly influences treatment outcome. Future research to devise new therapeutic strategies should therefore focus on these microenvironmental interactions, which could prove to be the Achilles' heel of cancer. ■

1. Heath, J. P. Epithelial cell migration in the intestine. *Cell Biol. Int.* **20**, 139–146 (1996).
2. Ferlay, J. *et al.* Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008. *Int. J. Cancer* **127**, 2893–2917 (2010).
3. Vermeulen, L., Sprick, M. R., Kemper, K., Stassi, G. & Medema, J. P. Cancer stem cells—old concepts, new insights. *Cell Death Differ.* **15**, 947–958 (2008).
4. Fearon, E. R. & Vogelstein, B. A genetic model for colorectal tumorigenesis. *Cell* **61**, 759–767 (1990).  
**This seminal study describes the sequential genetic events that occur during CRC progression.**
5. Vermeulen, L. *et al.* Wnt activity defines colon cancer stem cells and is regulated by the microenvironment. *Nature Cell Biol.* **12**, 468–476 (2010).  
**This study highlights the importance of the microenvironment in colon CSC maintenance.**
6. Barker, N. *et al.* Identification of stem cells in small intestine and colon by marker gene *Lgr5*. *Nature* **449**, 1003–1007 (2007).  
**This study identifies LGR5 as a marker of ISCs, and was the first study to use full genetic lineage-tracing to verify stem-cell potential.**
7. Breault, D. T. *et al.* Generation of *mTert*-GFP mice as a model to identify and study tissue progenitor cells. *Proc. Natl Acad. Sci. USA* **105**, 10420–10425 (2008).
8. Montgomery, R. K. *et al.* Mouse telomerase reverse transcriptase (*mTert*) expression marks slowly cycling intestinal stem cells. *Proc. Natl Acad. Sci. USA* **108**, 179–184 (2011).
9. Sangiorgi, E. & Capecchi, M. R. *Bmi1* is expressed *in vivo* in intestinal stem cells. *Nature Genet.* **40**, 915–920 (2008).
10. Zhu, L. *et al.* Prominin 1 marks intestinal stem cells that are susceptible to neoplastic transformation. *Nature* **457**, 603–607 (2009).
11. Cheng, H. & Leblond, C. P. Origin, differentiation and renewal of the four main epithelial cell types in the mouse small intestine. V. Unitarian Theory of the origin of the four epithelial cell types. *Am. J. Anat.* **141**, 537–561 (1974).
12. Lopez-Garcia, C., Klein, A. M., Simons, B. D. & Winton, D. J. Intestinal stem cell replacement follows a pattern of neutral drift. *Science* **330**, 822–825 (2010).
13. Snippert, H. J. *et al.* Intestinal crypt homeostasis results from neutral competition between symmetrically dividing *Lgr5* stem cells. *Cell* **143**, 134–144 (2010).
14. Crosnier, C., Stamatakis, D. & Lewis, J. Organizing cell renewal in the intestine: stem cells, signals and combinatorial control. *Nature Rev. Genet.* **7**, 349–359 (2006).
15. Hardwick, J. C., Kodach, L. L., Offerhaus, G. J. & van den Brink, G. R. Bone morphogenetic protein signalling in colorectal cancer. *Nature Rev. Cancer* **8**, 806–812 (2008).
16. Reya, T. & Clevers, H. Wnt signalling in stem cells and cancer. *Nature* **434**, 843–850 (2005).
17. Kim, K. A. *et al.* Mitogenic influence of human R-spondin1 on the intestinal epithelium. *Science* **309**, 1256–1259 (2005).
18. Pinto, D., Gregorieff, A., Begthel, H., & Clevers, H. Canonical Wnt signals are essential for homeostasis of the intestinal epithelium. *Genes Dev.* **17**, 1709–1713 (2003).
19. Korinek, V. *et al.* Depletion of epithelial stem-cell compartments in the small intestine of mice lacking Tcf-4. *Nature Genet.* **19**, 379–383 (1998).
20. Gregorieff, A. *et al.* Expression pattern of Wnt signaling components in the adult intestine. *Gastroenterology* **129**, 626–638 (2005).
21. van de Wetering, M. *et al.* The  $\beta$ -catenin/TCF-4 complex imposes a crypt progenitor phenotype on colorectal cancer cells. *Cell* **111**, 241–250 (2002).
22. Sato, T. *et al.* Paneth cells constitute the niche for *Lgr5* stem cells in intestinal crypts. *Nature* **469**, 415–418 (2010).
23. Fre, S. *et al.* Notch signals control the fate of immature progenitor cells in the intestine. *Nature* **435**, 964–968 (2005).
24. van Es, J. H. *et al.* Notch/ $\gamma$ -secretase inhibition turns proliferative cells in intestinal crypts and adenomas into goblet cells. *Nature* **435**, 959–963 (2005).
25. van Es, J. H., de Geest, N., van de Born, M., Clevers, H. & Hassan, B. A. Intestinal stem cells lacking the Math1 tumour suppressor are refractory to Notch inhibitors. *Nature Commun.* **1**, 18 (2010).
26. He, X. C. *et al.* BMP signaling inhibits intestinal stem cell self-renewal through suppression of Wnt- $\beta$ -catenin signaling. *Nature Genet.* **36**, 1117–1121 (2004).
27. Haramis, A. P. *et al.* De novo crypt formation and juvenile polyposis on BMP inhibition in mouse intestine. *Science* **303**, 1684–1686 (2004).
28. Auclair, B. A., Benoit, Y. D., Rivard, N., Mishina, Y. & Perreault, N. Bone morphogenetic protein signaling is essential for terminal differentiation of the intestinal secretory cell lineage. *Gastroenterology* **133**, 887–896 (2007).
29. Bijlsma, M. F., Spek, C. A. & Peppelenbosch, M. P. Hedgehog: an unusual signal transducer. *Bioessays* **26**, 387–394 (2004).
30. van Dop, W. A. *et al.* Depletion of the colonic epithelial precursor cell compartment upon conditional activation of the Hedgehog pathway. *Gastroenterology* **136**, 2195–2203 (2009).
31. van den Brink, G. R. *et al.* Indian Hedgehog is an antagonist of Wnt signaling in colonic epithelial cell differentiation. *Nature Genet.* **36**, 277–282 (2004).
32. van Dop, W. A. *et al.* Loss of Indian Hedgehog activates multiple aspects of a wound healing response in the mouse intestine. *Gastroenterology* **139**, 1665–1676 (2010).
33. Varnat, F., Zacchetti, G. & Altaba, A. Hedgehog pathway activity is required for the lethality and intestinal phenotypes of mice with hyperactive Wnt signaling. *Mech. Dev.* **127**, 73–81 (2010).
34. Ootani, A. *et al.* Sustained *in vitro* intestinal epithelial culture within a Wnt-dependent stem cell niche. *Nature Med.* **15**, 701–706 (2009).
35. Sato, T. *et al.* Single *Lgr5* stem cells build crypt-villus structures *in vitro* without a mesenchymal niche. *Nature* **459**, 262–265 (2009).  
**References 34 and 35 were the first to report an *in vitro* culture system to propagate intestinal crypts containing the various differentiated cell types, as well as ISCs.**
36. Cunningham, D. *et al.* Colorectal cancer. *Lancet* **375**, 1030–1047 (2010).
37. Tenesa, A. *et al.* Genome-wide association scan identifies a colorectal cancer susceptibility locus on 11q23 and replicates risk loci at 8q24 and 18q21. *Nature Genet.* **40**, 631–637 (2008).
38. Tomlinson, I. P. *et al.* A genome-wide association study identifies colorectal cancer susceptibility loci on chromosomes 10p14 and 8q23.3. *Nature Genet.* **40**, 623–630 (2008).
39. Markowitz, S. D. & Bertagnolli, M. M. Molecular origins of cancer: molecular basis of colorectal cancer. *N. Engl. J. Med.* **361**, 2449–2460 (2009).
40. Goss, K. H. & Groden, J. Biology of the adenomatous polyposis coli tumor suppressor. *J. Clin. Oncol.* **18**, 1967–1979 (2000).
41. Hanahan, D. & Weinberg, R. A. Hallmarks of cancer: the next generation. *Cell* **144**, 646–674 (2011).
42. Moser, A. R., Pitot, H. C. & Dove, W. F. A dominant mutation that predisposes to multiple intestinal neoplasia in the mouse. *Science* **247**, 322–324 (1990).
43. Barker, N. *et al.* Crypt stem cells as the cells-of-origin of intestinal cancer. *Nature* **457**, 608–611 (2009).
44. Huntly, B. J. *et al.* MOZ-TIF2, but not BCR-ABL, confers properties of leukemic stem cells to committed murine hematopoietic progenitors. *Cancer Cell* **6**, 587–596 (2004).
45. Shen, M. M., Wang, X., Economides, K. D., Walker, D. & Abate-Shen, C. Progenitor cells for the prostate epithelium: roles in development, regeneration, and cancer. *Cold Spring Harb. Symp. Quant. Biol.* **73**, 529–538 (2008).
46. Holland, E. C. *et al.* Combined activation of Ras and Akt in neural progenitors induces glioblastoma formation in mice. *Nature Genet.* **25**, 55–57 (2000).
47. Visvader, J. E. Cells of origin in cancer. *Nature* **469**, 314–322 (2011).  
**This excellent review discusses the evidence for a stem-cell origin of cancer.**
48. Takayama, T. *et al.* Aberrant crypt foci: detection, gene abnormalities, and clinical usefulness. *Clin. Gastroenterol. Hepatol.* **3**, S42–S45 (2005).
49. Novelli, M. R. *et al.* Polyclonal origin of colonic adenomas in an XO/XY patient with FAP. *Science* **272**, 1187–1190 (1996).
50. Albuquerque, C. *et al.* The 'just-right' signaling model: APC somatic mutations are selected based on a specific level of activation of the  $\beta$ -catenin signaling cascade. *Hum. Mol. Genet.* **11**, 1549–1560 (2002).
51. Crabtree, M. *et al.* Refining the relation between 'first hits' and 'second hits' at the APC locus: the 'loose fit' model and evidence for differences in somatic mutation spectra among patients. *Oncogene* **22**, 4257–4265 (2003).
52. Rubinfeld, B., Albert, I., Porfiri, E., Munemitsu, S. & Polakis, P. Loss of  $\beta$ -catenin regulation by the APC tumor suppressor protein correlates with loss of structure due to common somatic mutations of the gene. *Cancer Res.* **57**, 4624–4630 (1997).
53. Tanaka, T. *et al.* Dextran sodium sulfate strongly promotes colorectal carcinogenesis in *Apc*<sup>Min/+</sup> mice: inflammatory stimuli by dextran sodium sulfate results in development of multiple colonic neoplasms. *Int. J. Cancer* **118**, 25–34 (2006).
54. Su, L. K. *et al.* Multiple intestinal neoplasia caused by a mutation in the murine homolog of the APC gene. *Science* **256**, 668–670 (1992).
55. Taketo, M. M. & Edelmann, W. Mouse models of colon cancer. *Gastroenterology* **136**, 780–798 (2009).
56. Aoki, K., Tamai, Y., Horiike, S., Oshima, M. & Taketo, M. M. Colonic polyposis caused by mTOR-mediated chromosomal instability in *Apc*<sup>+/Δ716</sup> *Cdx2*<sup>+/−</sup> compound mutant mice. *Nature Genet.* **35**, 323–330 (2003).
57. Janssen, K. P. *et al.* APC and oncogenic KRAS are synergistic in enhancing Wnt signaling in intestinal tumor formation and progression. *Gastroenterology* **131**, 1096–1109 (2006).
58. Luo, F. *et al.* Mutated *K-ras*<sup>Asp12</sup> promotes tumorigenesis in *Apc*<sup>Min</sup> mice more in the large than the small intestines, with synergistic effects between K-ras and Wnt pathways. *Int. J. Exp. Pathol.* **90**, 558–574 (2009).
59. Marsh, V. *et al.* Epithelial Pten is dispensable for intestinal homeostasis but suppresses adenoma development and progression after *Apc* mutation. *Nature Genet.* **40**, 1436–1444 (2008).
60. Gould, K. A. *et al.* Genetic evaluation of candidate genes for the Mom1 modifier of intestinal neoplasia in mice. *Genetics* **144**, 1777–1785 (1996).
61. Takaku, K. *et al.* Suppression of intestinal polyposis in *Apc*<sup>Δ716</sup> knockout mice by an additional mutation in the cytosolic phospholipase A<sub>2</sub> gene. *J. Biol. Chem.* **275**, 34013–34016 (2000).
62. Takaku, K. *et al.* Intestinal tumorigenesis in compound mutant mice of both



- Dpc4 (Smad4)* and *Apc* genes. *Cell* **92**, 645–656 (1998).
63. Kitamura, T. *et al.* SMAD4-deficient intestinal tumors recruit CCR1<sup>+</sup> myeloid cells that promote invasion. *Nature Genet.* **39**, 467–475 (2007).
  64. Kodach, L. L. *et al.* The bone morphogenetic protein pathway is inactivated in the majority of sporadic colorectal cancers. *Gastroenterology* **134**, 1332–1341 (2008).
  65. Kim, B. G. *et al.* Smad4 signalling in T cells is required for suppression of gastrointestinal cancer. *Nature* **441**, 1015–1019 (2006).
  66. Katajisto, P. *et al.* LKB1 signaling in mesenchymal cells required for suppression of gastrointestinal polyposis. *Nature Genet.* **40**, 455–459 (2008).
  67. Itzkowitz, S. H. & Yio, X. Inflammation and cancer IV. Colorectal cancer in inflammatory bowel disease: the role of inflammation. *Am. J. Physiol. Gastrointest. Liver Physiol.* **287**, G7–G17 (2004).
  68. Chan, A. T., Ogino, S. & Fuchs, C. S. Aspirin use and survival after diagnosis of colorectal cancer. *J. Am. Med. Assoc.* **302**, 649–658 (2009).
  69. Oshima, M. *et al.* Suppression of intestinal polyposis in *Apc*<sup>Δ716</sup> knockout mice by inhibition of cyclooxygenase 2 (COX-2). *Cell* **87**, 803–809 (1996).
  70. Shao, J., Sheng, G. G., Mifflin, R. C., Powell, D. W. & Sheng, H. Roles of myofibroblasts in prostaglandin E2-stimulated intestinal epithelial proliferation and angiogenesis. *Cancer Res.* **66**, 846–855 (2006).
  71. Gounaris, E. *et al.* Mast cells are an essential hematopoietic component for polyp development. *Proc. Natl. Acad. Sci. USA* **104**, 19977–19982 (2007).
  72. Grivennikov, S. I., Greten, F. R. & Karin, M. Immunity, inflammation, and cancer. *Cell* **140**, 883–899 (2010).
  73. Greten, F. R. *et al.* IKK $\beta$  links inflammation and tumorigenesis in a mouse model of colitis-associated cancer. *Cell* **118**, 285–296 (2004).
  74. Lee, S. H. *et al.* ERK activation drives intestinal tumorigenesis in *Apc*<sup>min/+</sup> mice. *Nature Med.* **16**, 665–670 (2010).
  75. Wood, L. D. *et al.* The genomic landscapes of human breast and colorectal cancers. *Science* **318**, 1108–1113 (2007).
  76. Sottoriva, A. *et al.* Cancer stem cell tumor model reveals invasive morphology and increased phenotypical heterogeneity. *Cancer Res.* **70**, 46–56 (2010).
  77. Clarke, M. F. *et al.* Cancer stem cells—perspectives on current status and future directions: AACR workshop on cancer stem cells. *Cancer Res.* **66**, 9339–9344 (2006).
  78. Todaro, M. *et al.* Colon cancer stem cells dictate tumor growth and resist cell death by production of interleukin-4. *Cell Stem Cell* **1**, 389–402 (2007).
- This was the first study to provide evidence that colon CSCs are resistant to therapy.**
79. de Sousa, E. M., Vermeulen, L., Richel, D. J. & Medema, J. P. Targeting Wnt signaling in colon cancer stem cells. *Clin. Cancer Res.* **17**, 647–653 (2010).
  80. Todaro, M., Francipane, M. G., Medema, J. P. & Stassi, G. Colon cancer stem cells: promise of targeted therapy. *Gastroenterology* **138**, 2151–2162 (2010).
  81. Dalerba, P. *et al.* Phenotypic characterization of human colorectal cancer stem cells. *Proc. Natl. Acad. Sci. USA* **104**, 10158–10163 (2007).
  82. O'Brien, C. A., Pollett, A., Gallinger, S. & Dick, J. E. A human colon cancer cell capable of initiating tumour growth in immunodeficient mice. *Nature* **445**, 106–110 (2007).
  83. Ricci-Vitiani, L. *et al.* Identification and expansion of human colon-cancer-initiating cells. *Nature* **445**, 111–115 (2007).
- References 82 and 83 were the first studies to identify a CD133<sup>+</sup> subset of CRC cells with CSC properties.**
84. Vermeulen, L. *et al.* Single-cell cloning of colon cancer stem cells reveals a multi-lineage differentiation capacity. *Proc. Natl. Acad. Sci. USA* **105**, 13427–13432 (2008).
- This was the first study to show that primary colon CSCs can differentiate into multiple lineages depending on the extracellular signal input.**
85. Kemper, K. *et al.* The AC133 epitope, but not the CD133 protein, is lost upon cancer stem cell differentiation. *Cancer Res.* **70**, 719–729 (2010).
  86. Borovski, T., Vermeulen, L., Sprick, M. R. & Medema, J. P. One renegade cancer stem cell? *Cell Cycle* **8**, 803–808 (2009).
  87. Zucchi, I. *et al.* The properties of a mammary gland cancer stem cell. *Proc. Natl. Acad. Sci. USA* **104**, 10476–10481 (2007).
  88. Fodde, R. & Brabletz, T. Wnt/ $\beta$ -catenin signaling in cancer stemness and malignant behavior. *Curr. Opin. Cell Biol.* **19**, 150–158 (2007).
  89. Hoey, T. *et al.* DLL4 blockade inhibits tumor growth and reduces tumor-initiating cell frequency. *Cell Stem Cell* **5**, 168–177 (2009).
  90. Lombardo, Y. *et al.* Bone morphogenetic protein 4 induces differentiation of colorectal cancer stem cells and increases their response to chemotherapy in mice. *Gastroenterology* **140**, 297–309 (2011).
  91. Varnat, F., Siegl-Cachedenier, I., Malerba, M., Gervaz, P. & Altaba, A. Loss of WNT-TCF addiction and enhancement of HH-GLI1 signalling define the metastatic transition of human colon carcinomas. *EMBO Mol. Med.* **2**, 440–457 (2010).
  92. Mani, S. A. *et al.* The epithelial-mesenchymal transition generates cells with properties of stem cells. *Cell* **133**, 704–715 (2008).
  93. Potten, C. S., Gandara, R., Mahida, Y. R., Loeffler, M. & Wright, N. A. The stem cells of small intestinal crypts: where are they? *Cell Prolif.* **42**, 731–750 (2009).
  94. van der Flier, L. G. *et al.* Transcription factor achaete scute-like 2 controls intestinal stem cell fate. *Cell* **136**, 903–912 (2009).
  95. Schepers, A. G., Vries, R., van den, B. M., van de, W. M. & Clevers, H. Lgr5 intestinal stem cells have high telomerase activity and randomly segregate their chromosomes. *EMBO J.* **30**, 1104–1109 (2011).
  96. Ponder, B. A. *et al.* Derivation of mouse intestinal crypts from single progenitor cells. *Nature* **313**, 689–691 (1985).
  97. Loeffler, M., Birke, A., Winton, D. & Potten, C. Somatic mutation, monoclonality and stochastic models of stem cell organization in the intestinal crypt. *J. Theor. Biol.* **160**, 471–491 (1993).
  98. Huang, E. H. *et al.* Aldehyde dehydrogenase 1 is a marker for normal and malignant human colonic stem cells (SC) and tracks SC overpopulation during colon tumorigenesis. *Cancer Res.* **69**, 3382–3389 (2009).
  99. Campbell, F. *et al.* Post-irradiation somatic mutation and clonal stabilisation time in the human colon. *Gut* **39**, 569–573 (1996).
  100. Trumpp, A., Essers, M. & Wilson, A. Awakening dormant haematopoietic stem cells. *Nature Rev. Immunol.* **10**, 201–209 (2010).

**Acknowledgements** We would like to thank the members of the laboratory and G. van den Brink for critically reading the manuscript. Our work is sponsored by grants from the Netherlands Organization for Scientific Research (NWO; VICI program) and from the Dutch Cancer Society (UVA2009-4416).

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). Readers are welcome to comment on the online version of this article at [www.nature.com/nature](http://www.nature.com/nature). The authors declare no competing financial interests. Correspondence should be addressed to J.P.M. ([j.p.medema@amc.nl](mailto:j.p.medema@amc.nl)).

# Human nutrition, the gut microbiome and the immune system

Andrew L. Kau<sup>1\*</sup>, Philip P. Ahern<sup>1\*</sup>, Nicholas W. Griffin<sup>1</sup>, Andrew L. Goodman<sup>1†</sup> & Jeffrey I. Gordon<sup>1</sup>

**Marked changes in socio-economic status, cultural traditions, population growth and agriculture are affecting diets worldwide. Understanding how our diet and nutritional status influence the composition and dynamic operations of our gut microbial communities, and the innate and adaptive arms of our immune system, represents an area of scientific need, opportunity and challenge. The insights gleaned should help to address several pressing global health problems.**

Many recent reviews have described the known interactions between the innate and adaptive immune system and the tens of trillions of microbes that live in our gastrointestinal tracts (known as the gut microbiota). In this Perspective, we emphasize how the time is right and the need is great to understand better the relationships between diet, nutritional status, the immune system and microbial ecology in humans at different stages of life, living in distinct cultural and socio-economic settings.

This is a timely topic for many reasons. There is enormous pressure to devise ways to feed healthy foods to a human population whose size is predicted to expand to 9 billion by 2050. The solutions will have to address the challenges of developing sustainable forms of agriculture in the face of constrained land and water resources<sup>1</sup>. There is also a need to develop translational medicine pipelines to define more rigorously the nutritional value of foods that we consume and that we imagine creating in the future. These pipelines are required to evaluate health claims made about food ingredients. Increasing evidence shows that the nutritional value of food is influenced in part by the structure and operations of a consumer's gut microbial community, and that food, in turn, shapes the microbiota and its vast collection of microbial genes (the gut microbiome) (see, for example, refs 2 and 3). Therefore, to define the nutritional value of foods and our nutritional status better, we need to know more about our microbial differences and their origins, including how our lifestyles influence the assembly of gut microbial communities in children, and about the transmission of these communities within and across generations of a kinship<sup>4</sup>. We are learning how our gut microbial communities and immune systems co-evolve during our lifespans, and how components of the microbiota affect the immune system. We are also obtaining more information about how our overall metabolic phenotypes (metabotypes) reflect myriad functions encoded in our human genomes and gut microbiomes. These observations raise the question of how the metabolism of foods we consume by the gut microbial community affects our immune systems.

The link between infections that occur within and outside the gut and the development of nutritional deficiencies has been emphasized for many years. In turn, poor nutrition increases the risk of infection. Nonetheless, there is still a dearth of mechanistic information that explains these observations. Furthermore, only four years remain to achieve the United Nations eight Millennium Development Goals (<http://www.undp.org/mdg/>). Two of these goals relate to human nutrition: one seeks to eradicate extreme poverty and hunger, and

another aims to reduce the under-five mortality rate by two-thirds. Up to 1 billion people suffer from undernutrition of varying degrees, including 'silent' or asymptomatic malnutrition (<http://www.fao.org/publications/sofi/en/>), making this condition an enormous global health problem. Of the ~10 million children under the age of 5 who die every year, undernutrition contributes in some fashion to more than 50% of these deaths<sup>5</sup>. Sadly, children who survive periods of severe undernutrition can suffer long-term sequelae, including stunting and neurodevelopmental deficits<sup>6</sup>. Moreover, the effects of undernutrition can be felt across generations. Undernourished mothers suffer higher rates of morbidity and mortality, and are more likely to have low-birth-weight children, who have an increased risk of developing type 2 diabetes, hypertension, dyslipidaemia, cardiovascular pathology and obesity as adults<sup>7</sup>. A testable hypothesis is that the gut microbiota may contribute to the risk and pathogenesis of undernutrition through effects on nutrient metabolism and immune function (Fig. 1). Similarly, the experience of undernutrition in childhood could affect the development of metabolic capacities by this microbial 'organ' in ways that result in persistent metabolic dysfunction or inadequate function, thereby contributing to the sequelae of malnutrition. Finally, if we define malnutrition as the inadequate or excessive consumption of dietary ingredients leading to the development of disease, then we also need to consider the alarming epidemic of obesity that is sweeping the world and its relationship to the gut microbiome and the immune system.

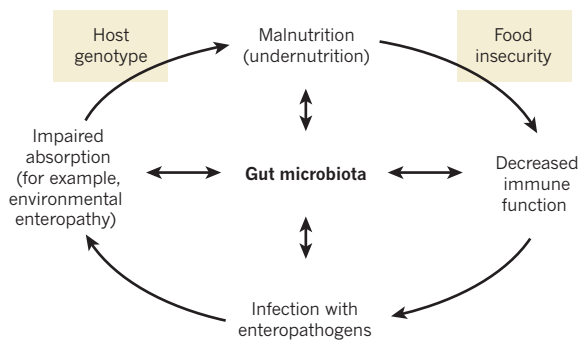
## The marriage of metagenomics and gnotobiotics

We believe that the 'marriage' of two approaches — one involving culture-independent (metagenomic) methods for describing the gut microbiota or microbiome and the other involving gnotobiotics (the rearing of animals under germ-free conditions, with or without subsequent exposure during postnatal life or adulthood to a microbial species or species consortium) — is a potentially powerful way to address several questions about the relationships between diet, nutritional status, the assembly and dynamic operations of gut microbial communities, and the nature of the interkingdom communications between the gut microbiota and the host (including host-microbial co-metabolism, and the co-evolution of the immune system<sup>3,8,9</sup>). Without dismissing caveats related to the use of gnotobiotic models (see later), in this Perspective we describe ways that may be useful for joining gnotobiotics and metagenomic methods to compare the functional properties of various types of gut microbial community, to explicitly test or generate hypotheses, and to develop new

<sup>1</sup>Center for Genome Sciences and Systems Biology, Washington University School of Medicine, St Louis, Missouri 63108, USA. <sup>†</sup>Present address: Section of Microbial Pathogenesis and Microbial Diversity Institute, Yale School of Medicine, New Haven, Connecticut 06536, USA.

\*These authors contributed equally to this work.





**Figure 1 | A schematic of the proposed relationships between the gut microbiota, the immune system and the diet, which underlie the development of malnutrition.** Undernutrition is associated with several defects in the innate and adaptive immune systems, which, in turn, are associated with increased predisposition to diarrhoeal illnesses. Recurrent (enteric) infections predispose to macronutrient and micronutrient deficiencies, as well as impaired intestinal mucosal barrier function<sup>77</sup>. These factors lead to a cycle of further susceptibility to infection and worsening nutritional status. A confounding problem is that vaccines designed to protect children from certain pathogens (including enteropathogens) show poor efficacy in areas of the world where poor nutrition is rampant<sup>74</sup>. One testable hypothesis is that the microbiota contributes to disease risk and pathogenesis. Diet shapes gut microbial community structure and function, and the microbiota adapts in ways that promote nutrient processing; the ability of the microbiota to process a given diet affects the nutrient and energetic value of that diet. The microbiota and immune systems co-evolve: malnutrition affects the innate and adaptive immune systems as well as the microbiota. The microbiota acts as a barrier to enteropathogen infection; this barrier function may be disrupted by malnutrition, as well as by perturbations in immune system function. The microbiota affects nutrient processing by the host, including the expression of host genes involved in nutrient transport and metabolism.

experimental (and computational) approaches that together inform the design, execution and interpretation of human studies.

### What is changing about what we eat?

Changes in dietary consumption patterns affect many aspects of human biology. To fully understand the determinants of nutritional status, we need to know what people are eating and how these diets are changing. Unfortunately, accurate information of this type is hard to obtain, and when available it generally covers a relatively limited time period. As a corollary, searchable databases that effectively integrate information obtained from the surveillance efforts of many international and national organizations (such as the World Health Organization, the UN Food and Agriculture Organization and the United States Department of Agriculture (USDA) Economic Research Service) are needed to monitor changing patterns of food consumption in different human populations. Analysis of USDA data that track the availability of more than 200 common food items between 1970 and 2000 shows that diets in the United States have changed in terms of both the overall caloric intake and the relative amounts of different food items (<http://www.ers.usda.gov/Data/FoodConsumption>). Linear regression of total caloric intake over time shows that the average number of kilocalories consumed per day increased markedly over this 30-year period ( $R^2 = 0.911$ ,  $P < 10^{-15}$ ). This is consistent with estimates from the US National Health and Nutrition Examination Survey (NHANES), which indicate that adult men and women increased their daily caloric intake by 6.9% and 21.7%, respectively, during the same period<sup>10</sup>. If total caloric intake is analogous to 'primary productivity' in macro-ecosystems, in which primary productivity is used as a proxy for available energy, then increasing the amount of energy input from the diet would be predicted to affect the number of microbial species living in the gut of a single host, as well as the magnitude of the compositional differences that exist between different hosts or even different regions of a single gut (see ref. 11 for discussions about the mechanisms underlying productivity–species

richness relationships in macro-ecosystems). Intriguingly, metagenomic studies of bacterial composition in the faecal microbiota of obese and lean twins living in the United States have shown that obesity is associated with decreased numbers of bacterial species<sup>4</sup>. Reductions in diversity could affect community function, resilience to various disturbances and the host immune system.

During the past 30 or so years, the North American diet has also shifted in terms of the relative contributions of different foods to total energy intake. Since 1970, two dietary 'epochs' can be distinguished based on the contribution of grains to overall calories (the mean increase in daily carbohydrate intake for men and women during this period was 62.4 g and 67.7 g, respectively<sup>10</sup>). The consumption of other food items has also changed: Spearman's rank correlations between food availability and time, followed by adjustments of  $P$  values to reflect false discovery rates, show that the representation of 177 out of 214 items tracked by the USDA has increased or decreased significantly in US diets since 1970. For example, Americans now eat less beef and more chicken, and corn-derived sweeteners have increased at the expense of cane and beet sugars. Furthermore, methods of food modification and preparation have changed. Comparable data are needed for other countries with distinct cultural traditions, including countries in which people are undergoing marked transformations in their socio-economic status and lifestyles.

We know from metagenomic studies of the human gut microbiota and microbiome that early postnatal environmental exposures have an important role in determining the overall phylogenetic structure of an adult human gut microbiota. The assembly of the microbiota towards an adult configuration occurs during the first three years of life<sup>12</sup>, and features of the organismal and gene content of gut communities are shared among family members and transmitted across generations of a kinship<sup>4</sup>. We also know that dietary habits influence the structure of the human genome. For example, populations that consume diets high in starch have a higher copy number of the salivary amylase gene (*AMY1*) than those consuming low-starch diets<sup>13</sup>. We know that these habits also affect the gut microbiome. A wonderful illustration of the latter point is provided by a microbial  $\beta$ -porphyranase in Japanese populations. *Zobellia galactanivorans* is a marine member of the Bacteroidetes that can process porphyran derived from marine red algae belonging to the *Porphyra* genus. Homologues of porphyranase genes from *Z. galactanivorans* are present in the human gut bacterium *Bacteroides plebeius* and are prominently represented in the microbiomes of Japanese but not North American citizens. This finding led to the suggestion that porphyranase genes from *Z. galactanivorans* or another related bacterium were acquired, perhaps through horizontal gene transfer, by a resident member of the microbiota of Japanese consumers of non-sterile food, and that this organism and gene were subsequently transmitted to others in Japanese society<sup>14</sup>. Together, these observations lead to the notion that systematic changes in overall dietary consumption patterns across a population might lead to changes in the microbiome, with consequences for host nutritional status and immune responses.

We also know, from work in gnotobiotic mice that have received human faecal microbial community transplants, that the relative abundances of different bacterial species and genes in the gut microbiota are highly sensitive to different foods<sup>3</sup>. Gnotobiotic mice containing defined collections of sequenced human gut symbionts or transplanted human faecal microbial communities could provide an approach for modelling the effects of different dietary epochs on the gut microbiota and on different facets of host biology. If the desired result is an account of the effects of individual food items or nutrients, then feeding the animals a series of defined diets, each with a different element removed or added, might be an appropriate strategy if the food ingredients for the epoch are known and available. If the focus is on the effects of overall differences in dietary habits within or between groups of humans, then diets should reflect the overall nutritional characteristics of the different groups and not merely be representative of a single individual.

Designing such diets requires detailed accounts of the identity and quantity of each food item consumed, ideally for a large number of people, as well as the methods used for food preparation. The US diet presents a rare opportunity for such an approach, because NHANES data sets (<http://www.cdc.gov/nchs/tutorials/Dietary/>) provide one-day dietary recall data at several time points since the early 1970s.

### Nutrient metabolism and the immune system

The nexus between nutrient metabolism and the immune system occurs at many levels, ranging from endocrine signalling to direct sensing of nutrients by immune cells.

Leptin signalling provides an example of these complex inter-relationships. Leptin regulates appetite and is a pleiotropic cytokine, maintaining thymic output and cellularity and promoting the dominance of T helper 1 ( $T_H1$ ) cells over  $T_H2$  cells<sup>15,16</sup> while inhibiting the proliferation of T regulatory ( $T_{reg}$ ) cells<sup>17</sup>. Low levels of leptin may account for the decreased cellular immunity associated with periods of nutrient deprivation<sup>16</sup>. Leptin also affects innate immune cells, ranging from the promotion of neutrophil activation and migration to the activation of monocytes and macrophages<sup>15</sup>. Elegant experiments using mice deficient in the leptin receptor in different cellular compartments showed a requirement for leptin signalling in intestinal epithelial cells to prevent severe disease after exposure to *Entamoeba histolytica*. Comparisons of *db/db* mice, which lack a functional leptin receptor, and their wild-type littermates demonstrated that leptin controls infectivity and prevents severe inflammatory destruction of the intestine, thereby affecting mortality<sup>18</sup>. These studies were extended to mice with engineered mutations in the leptin receptor that are found in human populations (Tyr1138Ser and Tyr985Leu, both of which disrupt signalling). These mutations rendered mice more susceptible to *E. histolytica* infection<sup>18</sup>. Leptin levels are significantly reduced in the sera of germ-free mice<sup>19</sup>. Moreover, obese, leptin-deficient (*ob/ob*) mice have marked differences in the taxonomic and genetic content of their gut microbial communities<sup>20</sup>. To our knowledge, the effects of leptin-receptor deficiency on the gut microbiota have not been reported. Nonetheless, leptin-receptor deficiency and *E. histolytica* pathogenesis provide a setting in which the intersections between the endocrine and immune systems, enteric infection and gut microbial ecology can be explored.

The ability to use macronutrients is essential for the generation and maintenance of a protective effector immune response. After stimulation through the T-cell receptor (TCR) and co-stimulation through CD28, the metabolic needs of T cells are met by a marked increase in the uptake and use of glucose and amino acids<sup>21,22</sup>. A deficiency in glucose uptake negatively affects numerous facets of T-cell function, with impairment of both proliferation and cytokine expression. Similarly, deficiencies in amino acids such as tryptophan, arginine, glutamine and cysteine reduce immune-cell activation. Furthermore, TCR stimulation in the absence of co-stimulation, which leads to T-cell anergy, has been linked to a failure to upregulate metabolic machinery associated with amino-acid and iron uptake<sup>21,22</sup>.

Short-chain fatty acids (SCFAs) provide one of the clearest examples of how nutrient processing by the microbiota and host diet combine to shape immune responses. SCFAs are end products of the microbial fermentation of macronutrients, most notably plant polysaccharides that cannot be digested by humans alone because our genomes do not encode the large repertoire of glycoside hydrolases and polysaccharide lyases needed to cleave the varied glycosidic linkages present in these glycans<sup>23</sup>. These missing enzymes are provided by the microbiome. The luminal concentration of intestinal SCFAs can be modified by the amount of fibre in the diet, which affects the composition of the microbiota<sup>24</sup>. In addition to acting as an energy source for the host, SCFAs exert notable effects on host immune responses. Low levels of butyrate modify the cytokine production profile of  $T_H$  cells<sup>25</sup> and promote intestinal epithelial barrier integrity<sup>26</sup>, which in turn can help to limit the exposure of the mucosal immune system to luminal microbes and prevent aberrant

inflammatory responses. Production of another SCFA, acetate, by the microbiota promotes the resolution of intestinal inflammation by the G-protein-coupled receptor GPR43 (ref. 27). A recent study highlighted the important role of acetate production in preventing infection with the enteropathogen *Escherichia coli* (0157:H7). This effect was linked to the ability of acetate to maintain gut epithelial barrier function<sup>28</sup>. Intriguingly, SCFAs may regulate the acetylation of lysine residues<sup>29</sup>, a covalent modification that affects proteins involved in a variety of signalling and metabolic processes. The role of this covalent modification in modulating the activity of proteins intimately involved in innate and adaptive immune responses needs to be explored further. It is tempting to speculate that the covalent or non-covalent linkage of products of microbial metabolism to host proteins produced within the intestine, or at extra-intestinal sites, will be discovered and shown to have important regulatory effects. These different protein modifications could represent a series of mechanisms by which the microbial community metabolite is 'imprinted' on the host.

If nutrients and derived metabolites reflect the functional activity of the microbiota, sensors of nutrient and metabolite availability can be considered akin to microbe-associated molecular patterns (MAMPs) that convey information about microbes to the host. Several families of innate receptors are involved in the recognition of MAMPs: these include Toll-like receptors (TLRs), inflammasomes, C-type lectins such as dectin-1, and RNA-sensing RIG-like helicases such as RIG-I and MDA5. The accompanying review by Maloy and Powrie (page 298) provides an overview of this area. We would like to emphasize that classical innate immune recognition pathways have evolved to survey the nutrient environment. TLR4 can sense the presence of free fatty acids<sup>30</sup>, whereas ATP is an important activator of the inflammasome<sup>31</sup>. Several other immune-cell-associated sensors couple information about the local nutrient or metabolite environment to the coordination of local immune responses. Examples are the serine/threonine kinase mammalian target of rapamycin (mTOR)<sup>32</sup>, double-stranded RNA-activated protein kinase (PKR)<sup>33</sup>, the aryl hydrocarbon receptor (AHR)<sup>34</sup>, and various nuclear hormone receptors such as the liver-X-receptor and the peroxisome-proliferator-activated receptors (PPAR- $\alpha$ , PPAR- $\beta$  and PPAR- $\gamma$ )<sup>35</sup> (Table 1 and Fig. 2). The mTOR pathway is an example of how energy availability affects immune responses. mTOR is activated by phosphatidylinositol-3-OH kinase and the serine/threonine kinase AKT, and is inhibited by AMP-activated protein kinase, which is a sensor of cellular energy resources. Genetic and pharmacological approaches (the latter using rapamycin) indicate that mTOR signalling affects both the innate and adaptive arms of the immune system — including maturation and effector activity of dendritic cells, inhibition of  $T_{reg}$ -cell development, promotion of the differentiation of  $T_H1$ ,  $T_H2$  and  $T_H17$  cells, regulation of CD8<sup>+</sup> T-cell trafficking and inhibition of memory T-cell formation<sup>32,36</sup>. PKR couples the presence of free fatty acids to immune activation, and has been implicated in the pathogenesis of obesity in mice fed a high-fat diet, including their development of immuno-inflammatory and insulin-resistant phenotypes<sup>33</sup> (see below). AHR is activated by several agonists, including kynurenine — a product of tryptophan metabolism by indoleamine-2,3-dioxygenase<sup>37,38</sup>. AHR modulates the differentiation of dendritic cells<sup>39</sup> and promotes  $T_H17$ -cell and  $T_{reg}$ -cell differentiation and effector activity<sup>40,41</sup>. Withdrawal of tryptophan and arginine controls immune responses<sup>42,43</sup>. The presence of an intact amino-acid starvation response in T cells is essential for the immunosuppressive activity of tryptophan depletion by indoleamine-2,3-dioxygenase<sup>44</sup>. This example illustrates how the ability of T cells to sense levels of a nutrient (tryptophan) in its local environment, rather than using the nutrient solely as a fuel source, is an important determinant of cell fate. If the assessment of local nutrient levels or metabolites is an important feature in the immune decision-making process, and if the products of microbial metabolism are previously unappreciated agonists or antagonists of immune-cell receptors, then an important challenge is to devise *in vitro* and *in vivo* models, including genetically manipulable



gnotobiotic animals (such as mice or zebrafish), to identify the range of metabolites produced by a microbiota (and host) as a function of different defined diets.

### The case for micronutrients

The intestinal microbiota can synthesize several vitamins involved in myriad aspects of microbial and host metabolism, including cobalamin (vitamin B<sub>12</sub>), pyridoxal phosphate (the active form of vitamin B<sub>6</sub>), which is involved in several enzymatic interconversions in amino-acid metabolism, pantothenic acid (vitamin B<sub>5</sub>), niacin (vitamin B<sub>3</sub>), biotin, tetrahydrofolate and vitamin K. In addition to vitamin B<sub>12</sub>, gut microbes produce a range of related molecules (corrinoids) with altered 'lower ligands', including analogues such as methyladenine and *p*-cresol. More than 80% of non-absorbed dietary vitamin B<sub>12</sub> is converted to these alternative corrinoids<sup>45,46</sup>. There is preliminary evidence to suggest that syntrophic relationships among members of the human microbiota, and the fitness of some taxa, may be based on the ability to generate, use or further transform various corrinoids<sup>46,47</sup>.

Folate and cobalamin produced by the gut microbiota could affect host DNA methylation patterns, whereas acetate produced by the microbial fermentation of polysaccharides could modify chromatin structure and gene transcription by histone acetylation. Thus, the inheritance of a mammalian genotype and intergenerational transmission of a microbiome — together with a complex dynamic in which the microbiome is viewed both as an epigenome and a modifier of the host epigenome during the postnatal period when host, host diet and microbial community co-evolve — could together shape human physiological phenotypes that are manifested during childhood or later in life.

Numerous observational studies indicate that deficiencies in vitamins A, D and E and zinc can adversely affect immune function, particularly T-cell responses. Although a considerable body of work exists detailing the myriad effects of vitamins A, D and E on host immune responses, so far there is little evidence for a role of the microbiota in the biosynthesis or metabolism of these vitamins. However, stimulation of dendritic cells through TLR2 increases the expression of host genes associated with generation of the immunoactive form of vitamin A (retinoic acid), and enteric infection has been linked to vitamin A deficiency<sup>48,49</sup>. Intriguingly, a recent study demonstrated that vitamin A deficiency leads to a complete loss of T<sub>H</sub>17 cells in the small intestine of specific pathogen-free mice and an associated significant reduction in the abundance of segmented filamentous bacteria (SFB)<sup>50</sup> — a member of the Clostridiaceae family that drives intestinal T<sub>H</sub>17 responses in mice<sup>51,52</sup>. Thus, vitamin A has the potential to modulate immune responses directly, by interacting with immune cells, or indirectly, by modulating the composition of the microbiota.

The microbiota also affects the absorption of key minerals. Perhaps the best characterized micronutrient in terms of its interaction with both the microbiota and the immune system is iron. Iron-deficient mice are resistant to the development of experimental autoimmune encephalomyelitis, and have reduced delayed type hypersensitivity (also known as type IV hypersensitivity) responses and lower levels of IgM and IgG. Iron deficiency also impairs innate immune responses, as it is required for the respiratory burst<sup>53</sup>. Likewise, iron is an essential micronutrient for bacteria. Given the low solubility of Fe<sup>3+</sup>, microbes have evolved the capacity to produce several high-affinity iron-binding siderophores. Microbes take up soluble Fe<sup>3+</sup>–siderophore complexes by several active transporters. Early studies in gnotobiotic animals showed a link between the gut microbiota and the development of iron deficiency. Germ-free but not conventionally raised rats become anaemic when fed a low-iron diet. The germ-free rats also show increased loss of iron in their faeces compared with their conventionally raised counterparts<sup>54</sup>. The iron balance that exists between host and microbiota is disturbed in a mouse model of Crohn's disease in which tumour-necrosis factor- $\alpha$  (TNF- $\alpha$ ) expression is dysregulated: oral (but not parenteral) iron supplementation in these animals causes a

shift in the gut microbial community composition, as defined by 16S ribosomal-RNA-based surveys, and exacerbates their ileitis<sup>55</sup>.

Metagenomic methods need to be applied to delineate further the role of the microbiota in micronutrient deficiencies. Several questions remain, such as how iron deficiency affects the configuration of the gut microbiota and microbiome, including the production of siderophores. Iron repletion could return the microbiota/microbiome to a normal, pre-deficient state; alternatively, there may be persistent structural and functional perturbations that require continued nutritional supplementation to correct. There may be particular configurations of the microbiota/microbiome that predispose the host to iron or other types of micronutrient deficiency. The iron content of mother's milk during postnatal life could also affect the assembly and metabolic operations of the microbiota. In principle, these questions can first be addressed in gnotobiotic mouse models, and also extended to macronutrient-deficient states.

### The microbiota and the immune system in obesity

Obesity, metabolic syndrome and diabetes illustrate the role that the diet–microbiota–immune axis has in shaping human systems biology. Although the marked increase in obesity worldwide can be linked to an ever-growing trend towards excessive caloric intake, the microbiota has also been implicated in obesity. Studies of a cohort of twins living in the United States indicate that the bacterial phylogenetic composition of the faecal microbiota and the representation of microbial genes involved in several aspects of nutrient metabolism in the faecal microbiome are different in lean versus obese twin pairs<sup>4</sup>. Different research groups using different primers to amplify bacterial 16S ribosomal RNA genes for culture-independent analyses of gut microbial ecology, and studying different human populations consuming different diets, have reported varying results concerning the bacterial phylogenetic composition of the microbiota in lean versus obese individuals<sup>56</sup>.

Evidence that a link exists between the microbiota and obesity comes from transplant experiments in gnotobiotic mice. Gut communities from leptin-deficient, *ob/ob*, mice or mice with diet-induced obesity induce a greater increase in adiposity when transferred to germ-free recipients than do communities from wild-type littermates or mice that have been given a healthy, calorically less-dense diet<sup>20,57</sup>. Germ-free mice are resistant to diet-induced obesity. Further studies have shown that the gut microbial community regulates the expression of genes that affect fatty-acid oxidation and fat deposition in adipocytes. For example, production of the secreted lipoprotein lipase (LPL) inhibitor angiopoietin-like protein 4 (ANGPTL4; also known as fasting-induced adipose factor) is suppressed by the microbiota: studies of germ-free and conventionalized wild-type and *Angptl4*<sup>−/−</sup> mice established that microbiota-mediated suppression of gut epithelial expression of this secreted LPL inhibitor results in increased LPL activity and fat storage in white adipose tissue<sup>19,58</sup>. Moreover, *Tlr5*-deficient mice have a gut microbiota with a distinct configuration from that encountered in wild-type littermate controls. When their gut microbiota is transplanted to wild-type, germ-free recipients, food intake is increased compared with recipients of microbiota transplants from wild-type mice: increased adiposity and hyperglycaemia ensue<sup>59</sup>. The mechanism underlying the increase in food consumption remains to be defined, although the authors of the study speculate that inflammatory signalling may desensitize insulin signalling in ways that lead to hyperphagia.

Obesity in mice and humans is associated with the infiltration of adipose tissue by macrophages, CD8<sup>+</sup> T cells<sup>60</sup> and CD4<sup>+</sup> T cells<sup>61,62</sup>, and with the expression of inflammatory cytokines and chemokines such as interleukin-6 (IL-6), IL-17, TNF- $\alpha$ , CC-chemokine ligand 2 (CCL2) and interferon- $\gamma$ <sup>60,62,63</sup>. By contrast, adipose tissue in lean mice is home to a population of immunosuppressive T<sub>reg</sub> cells that prevents inflammation<sup>64</sup>. Mice deficient in CC-chemokine receptor 2 (*Ccr2*) and with obesity induced by consumption of a high-fat diet have reduced macrophage infiltration of the adipose tissue and improved glucose tolerance relative to *Ccr2*-sufficient controls<sup>60</sup>, highlighting the role of factors in recruiting

**Table 1 | Metabolite sensors associated with immune cells**

Sensor	Agonist	Immune response affected
mTOR	S1P	Inhibits T <sub>reg</sub> -cell differentiation and maintenance <sup>88</sup>
	Leptin	Promotes T <sub>H</sub> 1-cell differentiation <sup>16</sup>
	Leptin	Inhibits T <sub>reg</sub> -cell proliferation and function <sup>17,89</sup>
AHR	6-formylindolo[3,2-b]carbazole	T <sub>H</sub> 17-cell differentiation and IL-22 production by T <sub>H</sub> 17 cells <sup>40,41</sup>
	2,3,7,8-tetrachlorodibenzo-p-dioxin	Promotes T <sub>reg</sub> -cell induction <sup>40</sup>
	Kynurenine	Promotes T <sub>reg</sub> -cell induction <sup>38</sup>
PKR	Free fatty acids; palmitic acid	Promotes insulin resistance through inhibitory phosphorylation of IRS-1 (ref. 33)
RAR-RXR	Retinoic acid	Promotes intestinal T-cell homing <sup>90</sup> ; promotes T <sub>reg</sub> -cell generation <sup>91</sup> ; promotes T-cell proliferation <sup>92</sup> ; promotes T <sub>H</sub> 2-cell differentiation over T <sub>H</sub> 1 cells <sup>93</sup>
VDR-RXR	1,25(OH) <sub>2</sub> vitamin D <sub>3</sub>	Inhibits lymphocyte proliferation <sup>94</sup> ; inhibits interferon- $\gamma$ , IL-17 and IL-2 expression <sup>95</sup> ; promotes emergence of T <sub>reg</sub> cells <sup>96</sup> ; drives antimicrobial peptide expression <sup>97</sup> ; promotes T-cell expression of CCR10 (ref. 98)
GPR120	$\omega$ -3 Fatty acids	Inhibits inflammatory cytokine production and chemotaxis by macrophages <sup>99</sup>
GPR43	Acetate	Promotes resolution of intestinal inflammation <sup>27</sup>
P2X receptors	ATP	Promotes T <sub>H</sub> 17-cell generation <sup>100</sup>

AHR, aryl hydrocarbon receptor; mTOR, mammalian target of rapamycin; PKR, double-stranded RNA-dependent protein kinase; RAR, retinoic acid receptor; RXR, retinoid X receptor; S1P, sphingosine-1-phosphate; VDR, vitamin D receptor.

inflammatory immune cells and their associated pro-inflammatory products in the pathogenesis of metabolic abnormalities associated with obesity. Blockade of TNF- $\alpha$ <sup>65</sup> or expanding T<sub>reg</sub> cells using anti-CD3 monoclonal antibody<sup>62</sup> prevents the onset of obesity-associated insulin resistance in a mouse model of diet-induced obesity.

Inflammation drives the development of insulin resistance through the phosphorylation of insulin receptor 1 by TNF- $\alpha$  activation of c-Jun amino-terminal protein kinase 1 (JNK1), and perhaps inhibitor of nuclear factor- $\kappa$ B kinase- $\beta$  (IKK- $\beta$ ), protein kinase C and mTOR. Whereas signalling by the adaptor protein MyD88 promotes the development of type 1 diabetes in pathogen-free NOD (non-obese diabetic) mice, germ-free *Myd88*<sup>-/-</sup> NOD animals are susceptible to this disorder<sup>66</sup>. These findings suggest that particular intestinal microbial configurations can promote or prevent inflammatory immune responses that drive metabolic dysfunction.

Mice fed a high-fat diet have increased serum levels of lipopolysaccharide<sup>67</sup>. Furthermore, genetically obese mice that are deficient in leptin or its receptor have reduced intestinal barrier function<sup>68</sup>. As noted earlier, SCFAs produced by microbial fermentation affect the barrier. Thus, it will be important to assess whether obese humans show similar reductions in barrier function. A high-fat diet alters the structure of the intestinal microbiota, potentially leading to a reduction in gut barrier integrity. The enhanced translocation of microbes and/or their antigens may result in increased microbial antigen load at extra-intestinal sites, enhanced immune stimulation and the development of insulin resistance. Furthermore, nutrients are known to activate inflammatory arms of the immune system directly<sup>69</sup>. The capacity of the intestinal microbiota to shape immune responses outside the intestine is well documented. Studies have highlighted the ability of the microbiota and specifically SFB to support the development of autoimmune arthritis<sup>70</sup>

and experimental autoimmune encephalomyelitis<sup>71</sup>, both of which have been linked to excessive T<sub>H</sub>17 responses.

Unfortunately, the spatial relationships between members of the microbiota and their proximity to elements of the gut-associated immune system in healthy individuals or individuals with mucosal barrier dysfunction are not well understood. Gnotobiotic mouse models of obesity may help to provide important insights about the biogeography of microbial communities along the length and width of the gut, including whether microbial consortia occupy ectopic sites that could affect the development and perpetuation of barrier dysfunction (such as in the crypts of Lieberkühn, where multipotential gut stem cells reside — as described by Medema and Vermeulen (page 318)). Newer methods, such as combinatorial labelling and spectral imaging fluorescence *in situ* hybridization (CLASI-FISH)<sup>72</sup>, offer a great deal of promise for characterizing the spatial features of microbe–microbe and microbe–host cell interactions in the gut mucosa, especially if they are applied to gnotobiotic models composed of defined collections of sequenced microbes.

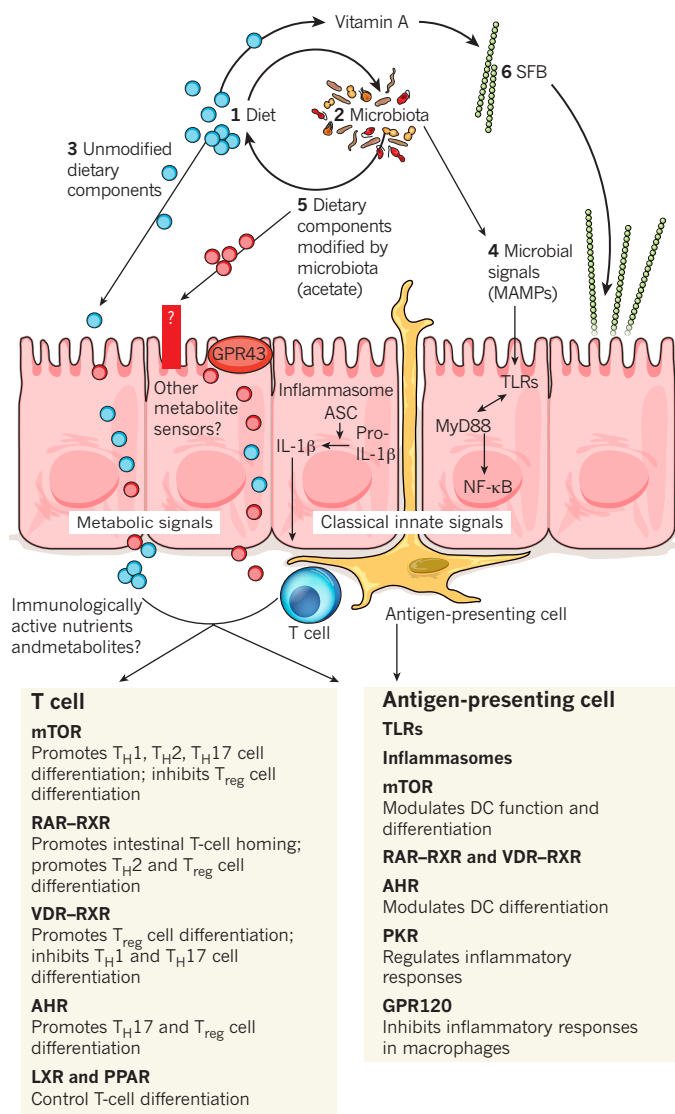
### Undernutrition and environmental enteropathy

Undernutrition can have many clinical manifestations ranging from mild, asymptomatic micronutrient deficiencies to severe, life-threatening conditions such as kwashiorkor and marasmus. Estimates indicate that the implementation of current ‘best practice’ interventions — including increasing the time of breastfeeding, supplementing diets with zinc and vitamins, hygiene measures such as improving hand washing, and optimizing the treatment of acute severe malnutrition — could reduce mortality during the first three years of life by only 25%, even if there is almost perfect compliance<sup>5</sup>. Several environmental and genetic factors have long been postulated to influence the development of moderate to severe forms of malnutrition<sup>73</sup>, but the underlying mechanisms remain poorly defined. Food availability, although a major factor, is not the only contributor. For example, in Malawi, the concordance for severe malnutrition between twins within the same household who are fed similar diets is only 50% (M. Manary, personal communication). This observation raises several questions. Do different configurations of the gut microbiota predispose one co-twin to kwashiorkor or marasmus? The effect of nutrient deficiency, in either the mother or her child, on the configuration of the microbiota and microbiome in the developing gut is not clear. It is possible that nutrient deficiency in the mother affects the assembly of the microbiota by causing changes in the mother’s gut microbiota or in the nutrient and immune content of her breast milk. Both the microbiota and milk are transmitted to the infant, yet we have much to learn about how the biochemical and immunological features of breast milk change, and how breast milk and the infant microbiota ‘co-evolve’ during the suckling period when a mother is healthy or malnourished (see below).

Identifying how malnutrition affects the gut’s microbiome may prove to be very important for improving many associated clinical disorders. Malnutrition could delay the maturation of the gut’s microbial metabolic organ or skew it towards a different and persistent configuration that lacks the necessary functions for health or increases the risk of diseases, including immuno-inflammatory disorders. Nutrient repletion may return the microbiota/microbiome to a ‘normal’ pre-deficient state; alternatively, structural and functional perturbations may persist, which require continued nutritional supplementation to correct. There may be microbiome configurations that correlate with vaccine responsiveness<sup>74</sup>.

Studies of severe forms of malnutrition indicate that these patients often have many characteristics of environmental enteropathy<sup>75</sup>. Also known as tropical sprue or tropical enteropathy, environmental enteropathy is a poorly characterized chronic inflammatory disease that mainly affects the small intestine. The disorder afflicts individuals who reside in areas with poor sanitation and who have high exposure to faecal-contaminated water and food. As an example, Peace Corps volunteers returning to the United States from such areas would report a history of diarrhoeal disease and have signs and symptoms of chronic





**Figure 2 | Metabolite sensors that help to coordinate immune responses.** Immune-cell-associated sensors use information about the local nutrient or metabolite milieu to organize local immune responses. (1) The dietary intake of macronutrients and micronutrients shapes microbial community structure (2), which, in turn, changes the nutritional value of the consumed food. (3) Unmodified dietary components are absorbed in the intestine, where they can interact with immune cells. (4) Microbial signals in the form of microbe-associated molecular patterns (MAMPs) modify local mucosal immune responses through innate signalling pathways such as the inflammasome or TLRs. Inflammasomes recruit the adaptor protein apoptosis-associated speck-like protein containing a CARD (ASC), which promotes binding of caspase, which in turn, cleaves pro-IL-1 $\beta$  to IL-1 $\beta$ . (5) Microbe-modified dietary components (such as acetate produced by the fermentation of polysaccharides) provide signals by which the immune system can monitor the metabolic activities of the microbiota. (6) Vitamin A can modify the representation of segmented filamentous bacteria (SFB) in the mouse gut microbiota, and is an example of a micronutrient directly modifying intestinal microbial ecology. SFB induce the differentiation of  $T_H17$  cells. DC, dendritic cell; LXR, liver-X-receptor; RAR, retinoic acid receptor; RXR, retinoid X receptor; VDR, vitamin D receptor.

malabsorption and nutritional deficiencies<sup>76</sup>. The malabsorption associated with environmental enteropathy is often subtle in children, manifesting itself clinically only as stunting due to chronic undernutrition<sup>76</sup>. The breakdown in intestinal mucosal barrier function in this disorder can lead to increased susceptibility to enteropathogen infections. Recurrent infections predispose individuals to nutritional

deficiencies and further compromise barrier function, leading to a vicious cycle of further susceptibility to infection and worsening nutritional status<sup>77</sup>.

Efforts to break this cycle have focused on vaccines that could prevent infection. However, there is significant heterogeneity in the responses to vaccination between children living in highly Westernized societies and children living in certain developing countries. Oral rotavirus vaccine elicits responses in more than 95% of children living in Westernized societies but in only 49% in Malawi<sup>78</sup>. Lower oral polio vaccine efficacy has been reported in populations with greater enteric disease burden<sup>79</sup>. Studies in Chilean children have demonstrated a negative correlation between oral cholera vaccine responses and small bowel bacterial overgrowth<sup>80</sup>. In addition, patients with coeliac disease, which shares phenotypic features with environmental enteropathy, can have a blunted response to parenteral hepatitis B vaccination, but only when their disease is active<sup>81</sup>.

Traditionally, the most definitive test for environmental enteropathy has been small intestinal biopsy. Biopsies typically show reductions in small intestinal villus height, increased numbers of intraepithelial lymphocytes, and increased infiltration of the underlying lamina propria by T cells with a predominant  $T_H1$  phenotype<sup>75</sup>. Some of these features are found in patients with coeliac disease, in which a luminal antigen (gliadin) drives a T-cell response that, in turn, results in epithelial destruction, reduced absorptive surface area and malabsorption<sup>76</sup>. Unlike coeliac disease, the antigens that drive the host immune response in environmental enteropathy are unknown, but there may be an association with certain HLA alleles, such as HLA-Aw31 (ref. 82).

The pathological events that lead to the development of environmental enteropathy are poorly understood, in part because of the absence of a robust set of readily assayed biomarkers that would improve the ability to diagnose, classify and potentially subcategorize individuals that show the broadly defined clinical manifestations that define this disorder. Epidemiological data showing a strong association of environmental enteropathy in areas with poor sanitation, as well as the occasional epidemic spread of the disease and its responsiveness to antibiotic treatment, reinforce the long-standing belief that there is an 'infectious' aetiology. Although cultures of jejunal aspirates from individuals with environmental enteropathy have suggested contamination of the proximal small bowel by aerotolerant Gram-negative bacteria<sup>83</sup>, no single pathogen or set of pathogens has been identified in the gut microbiota of most affected individuals. There is a distinct possibility that this enteropathy is not the result of a single pathogen but rather the result of colonization with microbial consortia that are inflammogenic in the context of a susceptible host. In fact, what constitutes a normal immune repertoire in a healthy gut probably varies considerably depending on environmental exposures and the configuration of a microbiota. Moreover, most metagenomic studies of the microbiota have focused on members of the superkingdom Bacteria, which dominate these communities. Other tools need to be developed so that they can be extended to viral, archaeal and eukaryotic components. The latter group includes parasites that compete for nutrients within the intestines of infected individuals. Parasites can interact directly with bacterial members of the microbiota during their life cycle in ways that promote hatching of parasite eggs, and can shape immune function through factors such as excretory-secretory products, which have been shown to modulate cytokine production, basophil degranulation and immune-cell recruitment and to interfere with TLR signalling<sup>84</sup>.

It seems reasonable to posit that individuals living in regions with high oral exposures to faecal-contaminated water and foods, and/or with a eukaryotic component of their gut community that includes parasites, will have gut-associated immune systems with significantly different structural and functional configurations than those without these exposures. In this sense, including the term environmental together with enteropathy is logical and emphasizes the need to place a host's immune and gut microbiome phenotypes in the context of their various exposures.

The representation and expression of microbiome genes in the gut communities of affected individuals compared with healthy controls may correlate with environmental enteropathy. Comparative metagenomic studies could thus provide important new diagnostic tools in the form of microbial taxa and microbiome gene functions. In addition, they could provide pathophysiological insight about relationships between host diet, enteropathogen representation in the microbiota, and microbiome gene composition and expression (including expressed metabolic functions). A major challenge will be to correlate this data with the results of quantitative phenotyping of the innate and adaptive immune systems of the human gut. This will require new and safe approaches for the sampling of immune system components, especially in the gut mucosa. Similarly, as noted earlier, the spatial relationships between members of the microbiota, as well as their proximity to elements of the gut-associated immune system in healthy individuals or in individuals with mucosal barrier dysfunction, is not well understood.

### Microbiota assembly and breast milk

Breast milk is known to protect newborns from infection, in part because of the copious quantity of maternally generated antibodies that it contains. Although these antibodies have specificity for components of the microbiota, the microbial targets are not well defined for given maternal–infant dyads, or as a function of time after delivery. In addition to antibodies, breast milk contains other immunoactive compounds, including cytokines such as IL-10, growth factors such as epidermal growth factor and antimicrobial enzymes such as lysozyme. The effect of maternal nutritional status on the glycan, protein, lipid and cytokine landscape of breast milk needs to be defined further. This analysis should have a temporal axis that explores co-evolution of the immunological and nutrient properties of mother's milk and the postnatal assembly and maturation of the infant gut microbiota and of the innate and adaptive immune systems. Important feedback systems may be revealed. Similarly, knowledge of the vaginal and cutaneous microbiota of mothers before and after birth, as a function of their nutritional status, could be very informative. For example, common configurations of microbial communities that occupy these body habitats could correlate with the development of environmental enteropathy in mothers and their offspring.

### Personalized gnotobiotics and culture collections

As noted above, studies have demonstrated the ability of intestinal microbial communities to reshape themselves rapidly in response to changes in diet. These observations raise the question of whether and how malnourished states affect (1) the spatial/functional organization of the microbiota and the niches (professions) of its component members; (2) the capacity of the community to respond to changes in diet; (3) the ability of components of the microbiota to forage adaptively on host-derived mucosal substrates; and (4) the physical and functional interactions that occur between the changing microbial communities and the intestinal epithelial barrier (including its overlying mucus layer). One way of developing the experimental and computational tools and concepts needed to examine these challenging questions in humans is to turn to gnotobiotic mice that have been 'humanized' by the transplantation of gut communities from human donors with distinct physiological phenotypes, and to feed these mice diets that are representative of those of the microbiota donor.

### Personalized gnotobiotic mouse models

We have used metagenomic methods to show that gut (faecal) communities can be efficiently transplanted into germ-free mice, and the mice can then be fed diets that resemble those consumed by the human microbiota donors, or diets with ingredients that are deliberately manipulated<sup>3,85</sup>. Transplanted human gut microbial communities can be transmitted from gnotobiotic mothers to their pups. In principle, mice humanized with microbiota from individuals

residing in different regions of the world, and given diets that are representative of those cultural traditions, can provide proof-of-principle global 'clinical trials' of the nutritional value of foods and their effect on the microbiota and the immune system.

Transplantation of a human faecal microbiota into germ-free mice can be viewed as capturing an individual's microbial community at a moment in time and replicating it in several recipient gut ecosystems. The humanized mice can be followed over time under highly controlled conditions in which potentially confounding variables can be constrained in ways that are not achievable in human studies. This type of personalized gnotobiotics also provides an opportunity to determine the degree to which human phenotypes can be transmitted via the gut microbiota as a function of diet. Moreover, the documented responses of microbial lineages and genes encoding metabolic pathways in the transplanted, replicated communities may provide mechanistic insight into differences in the adaptations of healthy versus diseased gut microbiomes (and the host immune system) to changes in diets, plus new biomarkers of nutritional status and the effect of various therapeutic interventions, including those based on dietary manipulations. Putative microbial biomarkers obtained from studies of these mice can, in turn, be used to query data sets generated directly from the human donor(s).

Despite the potential power of using humanized mice to study interactions between the host immune and metabolic systems and the intestinal microbiota under highly controlled conditions, this approach has caveats. Recent work on  $T_H17$  responses suggests that unlike the mouse microbiota, which contains SFB, a faecal microbiota from a human donor is not sufficient to drive the expression of immune-related genes in the small intestine of previously germ-free mice<sup>52</sup>. This raises the possibility that humanization may not fully recapitulate the capacity of a mouse microbiota to mature the intestinal immune system in mice. However, earlier studies on the effects of human microbiota on the mouse immune system showed that the ability of *E. coli* heat-labile enterotoxin to break oral tolerance to ovalbumin in germ-free mice can be inhibited by transplantation of either a human or a mouse microbiota during the neonatal period<sup>86</sup>. Furthermore, a single component of a human gut symbiont, the polysaccharide A component of *Bacteroides fragilis*, can mature components of the  $CD4^+$  T-cell response in mice<sup>87</sup>. Finally, we have observed a similar increase in the frequency of TCR- $\beta^+$  cells among lymphocytes in the mesenteric lymph nodes of gnotobiotic recipients of a human or mouse microbiota, compared with germ-free controls (P.P.A., V. K. Ridaura and J.I.G., unpublished observations). This suggests that although not all components of the mouse immune system will be matured by a human gut microbiota, the immune system is not likely to remain ignorant of these communities. In addition, any differences detected in direct comparisons of the effects of two different human gut communities may represent responses relevant to the human immune system.

### Personalized bacterial culture collections

We have recently shown that the human faecal microbiota consists largely of bacteria that can readily be cultured<sup>3</sup>. Metagenomic analysis indicates that most of the predicted functions of a human's microbiome are represented in its cultured members. In gnotobiotic mice, both complete and cultured communities have similar properties and responses to dietary manipulations. By changing the diet of the host, the community of cultured microbes can be shaped so that it becomes enriched for taxa suited to that diet. These culture collections of anaerobes can be clonally arrayed in multiwell formats: this means that personalized, taxonomically defined culture collections can be created from donors representing different human populations and physiological phenotypes in which the cultured microbes have co-evolved within a single human donor's gut habitat.

Together, these advances yield a translational medicine pipeline for examining the interactions between food and food ingredients, the microbiota, the immune system and health. The goals for such a human translational medicine pipeline are to identify individuals



with interesting phenotypes, to assess the transmissibility of their phenotypes by human microbiota transplants into gnotobiotic animals, to select candidate disease-modifying taxa (retrieved from clonally arrayed, taxonomically defined personal bacterial culture collections), to sequence selected taxa and to reunite them in various combinations in gnotobiotic mice as defined model gut communities. The interactions of disease-modifying taxa with one another, their effects on host biology, and how these effects are influenced by diet can be further explored using methods such as high-throughput complementary DNA sequencing (RNA-Seq), mass-spectrometry-based proteomics and metabolomics, multilabel fluorescence *in situ* hybridization (for biogeographical studies of the microbiota), whole-genome transposon mutagenesis (to identify fitness factors for microbes under various dietary contexts<sup>46</sup>), and immune profiling and other measurements of mucosal barrier function. Knowing the degree to which tractable bacterial taxa can influence host physiology, and how dietary components can be used to affect specific organisms in the microbiota<sup>3,85</sup> in ways that provide benefit to the host may be useful for discovering new generations of probiotics and prebiotics.

### Looking ahead

With massive prospective national surveys planned and being implemented — such as the National Institutes of Health National Children's Study, which will follow a representative sample of 100,000 children from before birth to age 21 — the time is right for an initiative to evaluate the relationships between our diets, nutritional status, microbiomes and immune systems. Many components could constitute this initiative. We can readily foresee several of these.

### Dietary databases

As noted earlier, there is a need to create more and improved databases for monitoring changing patterns of food consumption, in which the surveillance efforts of several organizations can be integrated. This tool and other interdisciplinary approaches could be used to define a set of study populations that represents established and emerging food consumption patterns in distinct cultural and socio-economic settings. An emphasis could be placed on comparing humans living in Westernized societies with those living in developing countries that are undergoing marked transitions in lifestyles and cultural traditions. New, reliable, cost-effective and generalized methods will be needed to acquire quantitative data about the diets consumed by individuals in these study populations, and the resultant data will need to be deposited in searchable databases using defined annotation standards. Moreover, guidelines related to the ethical and legal aspects of human-subject research involving observational and interventional nutritional studies of pregnant women and their offspring need to be further developed.

### New biomarkers of nutritional status

Readily procured human biospecimens could be used together with high-throughput, targeted and non-targeted (quantitative) profiling of metabolites in comprehensive time-series studies to define the relationship between diet, nutritional status and microbiome configuration in healthy individuals at various stages of life (for example, in women before, during and after pregnancy, and in their children during the first five years after birth). This could be accompanied by studies of malnourished individuals before, during and after well-justified, defined nutritional interventions. In addition to these data, genomes (genotypes), epigenomes and microbiomes could be characterized in these study cohorts together with a variety of clinical parameters (such as vaccine responses) and environmental parameters (such as water sanitation). The resultant data sets would be deposited in annotated searchable databases. A translational medicine pipeline that includes relevant cellular and animal models would help to guide the design and interpretation of these human studies.

### Quantitative phenotyping of the immune system

As noted earlier, a major challenge is to obtain cellular and molecular biomarkers for quantitative profiling of the innate and adaptive immune systems, including biomarkers of mucosa-associated barrier function. Given the small quantities of biomaterials available from certain body sites, this initiative should help to advance 'miniaturizing technology' for quantitative measurements of cells and biofluids. Non-invasive imaging-based biomarkers are also needed.

Goals include identifying new host and microbial biomarkers and mediators of nutritional status, determining the nutritional value of various foods, and characterizing the function of the human adaptive and innate immune systems (including mucosal barrier integrity and mucosal immunity) and the dynamic operations of the microbiota. This information would be used for demonstration projects that rigorously define nutritional health and test preventive or therapeutic recommendations for micronutrient and macronutrient consumption, for example in pregnant women and infants/children, and their effect on the assembly and operations of the immune system. The microbiome component could also help to define a previously uncharacterized axis of human genetic evolution (our microbiome evolution), reflecting in part our changing dietary habits. It could also produce testable hypotheses about unappreciated aspects of the pathophysiology of Western diseases, and yield new microbiome-based strategies for disease prevention or treatment. ■

- Whitacre, P. T., Fagen, A. P., Husbands, J. L. & Sharples, F. E. *Implementing the New Biology: Decadal Challenges Linking Food, Energy, and the Environment* (National Research Council of The National Academies of Science, 2010).
  - Muegge, B. *et al.* Diet drives convergence in gut microbiome functions across mammalian phylogeny and within humans. *Science* **332**, 970–974 (2011).
  - Goodman, A. L. *et al.* Extensive personal human gut microbiota culture collections characterized and manipulated in gnotobiotic mice. *Proc. Natl Acad. Sci. USA* **108**, 6252–6257 (2011).
- This report highlights the use of gnotobiotic mice containing a transplanted human gut microbiome for studying the dynamic interactions between diet and the microbial community.**
- Turnbaugh, P. J. *et al.* A core gut microbiome in obese and lean twins. *Nature* **457**, 480–484 (2009).
  - Bryce, J., Bosch-Pinto, C., Shibuya, K. & Black, R. E. WHO estimates of the causes of death in children. *Lancet* **365**, 1147–1152 (2005).
  - Bhutta, Z. A. *et al.* What works? Interventions for maternal and child undernutrition and survival. *Lancet* **371**, 417–440 (2008).
  - Barker, D. J. Adult consequences of fetal growth restriction. *Clin. Obstet. Gynecol.* **49**, 270–283 (2006).
  - Wikoff, W. R. *et al.* Metabolomics analysis reveals large effects of gut microflora on mammalian blood metabolites. *Proc. Natl Acad. Sci. USA* **106**, 3698–3703 (2009).
  - Martin, F. P. *et al.* Probiotic modulation of symbiotic gut microbial–host metabolic interactions in a humanized microbiome mouse model. *Mol. Syst. Biol.* **4**, 157 (2008).
  - Wright, J. D., Kennedy-Stephenson, J., Wang, C. Y., McDowell, M. A. & Johnson, C. L. Trends in intake of energy and macronutrients — United States, 1971–2000. *MMWR Morb. Mortal. Wkly Rep.* **53**, 80–82 (2004).
  - Chase, J. M. Stochastic community assembly causes higher biodiversity in more productive environments. *Science* **328**, 1388–1391 (2010).
  - Koenig, J. E. *et al.* Succession of microbial consortia in the developing infant gut microbiome. *Proc. Natl Acad. Sci. USA* **108**, 4578–4585 (2011).
  - Perry, G. H. *et al.* Diet and the evolution of human amylase gene copy number variation. *Nature Genet.* **39**, 1256–1260 (2007).
  - Hehemann, J. H. *et al.* Transfer of carbohydrate-active enzymes from marine bacteria to Japanese gut microbiota. *Nature* **464**, 908–912 (2010).
  - La Cava, A. & Matarese, G. The weight of leptin in immunity. *Nature Rev. Immunol.* **4**, 371–379 (2004).
  - Lord, G. M. *et al.* Leptin modulates the T-cell immune response and reverses starvation-induced immunosuppression. *Nature* **394**, 897–901 (1998).
  - De Rosa, V. *et al.* A key role of leptin in the control of regulatory T cell proliferation. *Immunity* **26**, 241–255 (2007).
  - Guo, X. *et al.* Leptin signaling in intestinal epithelium mediates resistance to enteric infection by *Entamoeba histolytica*. *Mucosal Immunol.* **4**, 294–303 (2011).
- This study demonstrates the role of leptin-receptor signalling in protecting the intestinal epithelium against infection and damage by the enteropathogen *E. histolytica*.**
- Backhed, F. *et al.* The gut microbiota as an environmental factor that regulates fat storage. *Proc. Natl Acad. Sci. USA* **101**, 15718–15723 (2004).
  - Turnbaugh, P. J. *et al.* An obesity-associated gut microbiome with increased capacity for energy harvest. *Nature* **444**, 1027–1031 (2006).
  - Fox, C. J., Hammerman, P. S. & Thompson, C. B. Fuel feeds function: energy metabolism and the T-cell response. *Nature Rev. Immunol.* **5**, 844–852 (2005).

22. Michalek, R. D. & Rathmell, J. C. The metabolic life and times of a T-cell. *Immunity*. **Rev.** **236**, 190–202 (2010).
23. Qin, J. *et al.* A human gut microbial gene catalogue established by metagenomic sequencing. *Nature* **464**, 59–65 (2010).
24. Lupton, J. R. Microbial degradation products influence colon cancer risk: the butyrate controversy. *J. Nutr.* **134**, 479–482 (2004).
25. Bird, J. J. *et al.* Helper T cell differentiation is controlled by the cell cycle. *Immunity* **9**, 229–237 (1998).
26. Peng, L., He, Z., Chen, W., Holzman, I. R. & Lin, J. Effects of butyrate on intestinal barrier function in a Caco-2 cell monolayer model of intestinal barrier. *Pediatr. Res.* **61**, 37–41 (2007).
27. Maslowski, K. M. *et al.* Regulation of inflammatory responses by gut microbiota and chemoattractant receptor GPR43. *Nature* **461**, 1282–1286 (2009).
28. Fukuda, S. *et al.* Bifidobacteria can protect from enteropathogenic infection through production of acetate. *Nature* **469**, 543–547 (2011).
- References 27 and 28 demonstrate how microbiota-derived short-chain fatty acids help to modulate immune responses and susceptibility to enteropathogen invasion.**
29. Kim, G. W., Gocevski, G., Wu, C. J. & Yang, X. J. Dietary, metabolic, and potentially environmental modulation of the lysine acetylation machinery. *Int. J. Cell Biol.* **2010**, 632739 (2010).
30. Nguyen, M. T. *et al.* A subpopulation of macrophages infiltrates hypertrophic adipose tissue and is activated by free fatty acids via Toll-like receptors 2 and 4 and JNK-dependent pathways. *J. Biol. Chem.* **282**, 35279–35292 (2007).
31. Mariathasan, S. *et al.* Cryopyrin activates the inflammasome in response to toxins and ATP. *Nature* **440**, 228–232 (2006).
32. Thomson, A. W., Turnquist, H. R. & Raimondi, G. Immunoregulatory functions of mTOR inhibition. *Nature Rev. Immunol.* **9**, 324–337 (2009).
33. Nakamura, T. *et al.* Double-stranded RNA-dependent protein kinase links pathogen sensing with stress and metabolic homeostasis. *Cell* **140**, 338–348 (2010).
34. Stockinger, B. Beyond toxicity: aryl hydrocarbon receptor-mediated functions in the immune system. *J. Biol.* **8**, 61 (2009).
35. Glass, C. K. & Ogawa, S. Combinatorial roles of nuclear receptors in inflammation and immunity. *Nature Rev. Immunol.* **6**, 44–55 (2006).
36. Araki, K., Youngblood, B. & Ahmed, R. The role of mTOR in memory CD8 T-cell differentiation. *Immunity*. **Rev.** **235**, 234–243 (2010).
37. Esser, C., Rannug, A. & Stockinger, B. The aryl hydrocarbon receptor in immunity. *Trends Immunol.* **30**, 447–454 (2009).
38. Mezrich, J. D. *et al.* An interaction between kynurenine and the aryl hydrocarbon receptor can generate regulatory T cells. *J. Immunol.* **185**, 3190–3198 (2010).
39. Platzer, B. *et al.* Aryl hydrocarbon receptor activation inhibits *in vitro* differentiation of human monocytes and Langerhans dendritic cells. *J. Immunol.* **183**, 66–74 (2009).
40. Quintana, F. J. *et al.* Control of T<sub>reg</sub> and T<sub>H</sub>17 cell differentiation by the aryl hydrocarbon receptor. *Nature* **453**, 65–71 (2008).
41. Veldhoen, M. *et al.* The aryl hydrocarbon receptor links T<sub>H</sub>17-cell-mediated autoimmunity to environmental toxins. *Nature* **453**, 106–109 (2008).
42. Bronte, V. & Zanovello, P. Regulation of immune responses by L-arginine metabolism. *Nature Rev. Immunol.* **5**, 641–654 (2005).
43. Mellor, A. L. & Munn, D. H. IDO expression by dendritic cells: tolerance and tryptophan catabolism. *Nature Rev. Immunol.* **4**, 762–774 (2004).
44. Munn, D. H. *et al.* GCN2 kinase in T cells mediates proliferative arrest and anergy induction in response to indoleamine 2,3-dioxygenase. *Immunity* **22**, 633–642 (2005).
45. Allen, R. H. & Stabler, S. P. Identification and quantitation of cobalamin and cobalamin analogues in human feces. *Am. J. Clin. Nutr.* **87**, 1324–1335 (2008).
46. Goodman, A. L. *et al.* Identifying genetic determinants needed to establish a human gut symbiont in its habitat. *Cell Host Microbe* **6**, 279–289 (2009).
47. Anderson, P. J. *et al.* One pathway can incorporate either adenine or dimethylbenzimidazole as an  $\alpha$ -axial ligand of B12 cofactors in *Salmonella enterica*. *J. Bacteriol.* **190**, 1160–1171 (2008).
48. Curtale, F., Pokhrel, R. P., Tilden, R. L. & Higashi, G. Intestinal helminths and xerophthalmia in Nepal. A case control study. *J. Trop. Pediatr.* **41**, 334–337 (1995).
49. Sommer, A., Tarwotjo, I. & Katz, J. Increased risk of xerophthalmia following diarrhea and respiratory disease. *Am. J. Clin. Nutr.* **45**, 977–980 (1987).
50. Cha, H. R. *et al.* Downregulation of Th17 cells in the small intestine by disruption of gut flora in the absence of retinoic acid. *J. Immunol.* **184**, 6799–6806 (2010).
- This study shows how a single micronutrient, vitamin A, modulates host immune responses through its effects on the composition of the intestinal microbiota.**
51. Ivanov, I. I. *et al.* Induction of intestinal Th17 cells by segmented filamentous bacteria. *Cell* **139**, 485–498 (2009).
52. Gaboriau-Routhiau, V. *et al.* The key role of segmented filamentous bacteria in the coordinated maturation of gut helper T cell responses. *Immunity* **31**, 677–689 (2009).
- References 51 and 52 are seminal studies identifying a single member of the intestinal microbiota that drives the differentiation of intestinal T<sub>H</sub>17 cells.**
53. Schaible, U. E. & Kaufmann, S. H. Iron and microbial infection. *Nature Rev. Microbiol.* **2**, 946–953 (2004).
54. Reddy, B. S., Pleasants, J. R. & Wostmann, B. S. Effect of intestinal microflora on iron and zinc metabolism, and on activities of metalloenzymes in rats. *J. Nutr.* **102**, 101–107 (1972).
55. Werner, T. *et al.* Depletion of luminal iron alters the gut microbiota and prevents Crohn's disease-like ileitis. *Gut* **60**, 325–333 (2011).
56. Ley, R. E. Obesity and the human microbiome. *Curr. Opin. Gastroenterol.* **26**, 5–11 (2010).
57. Turnbaugh, P. J., Backhed, F., Fulton, L. & Gordon, J. I. Diet-induced obesity is linked to marked but reversible alterations in the mouse distal gut microbiome. *Cell Host Microbe* **3**, 213–223 (2008).
58. Mandard, S. *et al.* The fasting-induced adipose factor/angiopoietin-like protein 4 is physically associated with lipoproteins and governs plasma lipid levels and adiposity. *J. Biol. Chem.* **281**, 934–944 (2006).
59. Vijay-Kumar, M. *et al.* Metabolic syndrome and altered gut microbiota in mice lacking Toll-like receptor 5. *Science* **328**, 228–231 (2010).
- This paper links changes in the configuration of the intestinal microbiota in Tlr5-deficient mice to inflammation and development of metabolic syndrome.**
60. Gregor, M. F. & Hotamisligil, G. S. Inflammatory mechanisms in obesity. *Annu. Rev. Immunol.* **29**, 415–445 (2011).
61. Kintscher, U. *et al.* T-lymphocyte infiltration in visceral adipose tissue: a primary event in adipose tissue inflammation and the development of obesity-mediated insulin resistance. *Arterioscler. Thromb. Vasc. Biol.* **28**, 1304–1310 (2008).
62. Winer, S. *et al.* Normalization of obesity-associated insulin resistance through immunotherapy. *Nature Med.* **15**, 921–929 (2009).
63. Zuniga, L. A. *et al.* IL-17 regulates adipogenesis, glucose homeostasis, and obesity. *J. Immunol.* **185**, 6947–6959 (2010).
64. Feuerer, M. *et al.* Lean, but not obese, fat is enriched for a unique population of regulatory T cells that affect metabolic parameters. *Nature Med.* **15**, 930–939 (2009).
65. Uysal, K. T., Wiesbrock, S. M., Marino, M. W. & Hotamisligil, G. S. Protection from obesity-induced insulin resistance in mice lacking TNF- $\alpha$  function. *Nature* **389**, 610–614 (1997).
66. Wen, L. *et al.* Innate immunity and intestinal microbiota in the development of type 1 diabetes. *Nature* **455**, 1109–1113 (2008).
67. Cani, P. D. *et al.* Metabolic endotoxemia initiates obesity and insulin resistance. *Diabetes* **56**, 1761–1772 (2007).
68. Brun, P. *et al.* Increased intestinal permeability in obese mice: new evidence in the pathogenesis of nonalcoholic steatohepatitis. *Am. J. Physiol. Gastrointest. Liver Physiol.* **292**, G518–G525 (2007).
69. Shi, H. *et al.* TLR4 links innate immunity and fatty acid-induced insulin resistance. *J. Clin. Invest.* **116**, 3015–3025 (2006).
70. Wu, H. J. *et al.* Gut-residing segmented filamentous bacteria drive autoimmune arthritis via T helper 17 cells. *Immunity* **32**, 815–827 (2010).
71. Lee, Y. K., Menezes, J. S., Umesaki, Y. & Mazmanian, S. K. Proinflammatory T-cell responses to gut microbiota promote experimental autoimmune encephalomyelitis. *Proc. Natl Acad. Sci. USA* **108**, 4615–4622 (2011).
72. Valm, A. M. *et al.* Systems-level analysis of microbial community organization through combinatorial labeling and spectral imaging. *Proc. Natl Acad. Sci. USA* **108**, 4152–4157 (2011).
73. Golden, M. H. Oedematous malnutrition. *Br. Med. Bull.* **54**, 433–444 (1998).
74. Ferreira, R. B., Antunes, L. C. & Finlay, B. B. Should the human microbiome be considered when developing vaccines? *PLoS Pathogens* **6**, e1001190 (2010).
75. Campbell, D. I. *et al.* Chronic T cell-mediated enteropathy in rural west African children: relationship with nutritional status and small bowel function. *Pediatr. Res.* **54**, 306–311 (2003).
76. Humphrey, J. H. Child undernutrition, tropical enteropathy, toilets, and handwashing. *Lancet* **374**, 1032–1035 (2009).
- This is an excellent review of the relationship between environmental enteropathy and malnutrition.**
77. Guerrant, R. L., Oria, R. B., Moore, S. R., Oria, M. O. & Lima, A. A. Malnutrition as an enteric infectious disease with long-term effects on child development. *Nutr. Rev.* **66**, 487–505 (2008).
78. World Health Organization. Meeting of the immunization Strategic Advisory Group of Experts, April 2009 — conclusions and recommendations. *Wkly Epidemiol. Rec.* **84**, 220–236 (2009).
79. Grassly, N. C. *et al.* Mucosal immunity after vaccination with monovalent and trivalent oral poliovirus vaccine in India. *J. Infect. Dis.* **200**, 794–801 (2009).
80. Lagos, R. *et al.* Effect of small bowel bacterial overgrowth on the immunogenicity of single-dose live oral cholera vaccine CVD 103-HgR. *J. Infect. Dis.* **180**, 1709–1712 (1999).
81. Nemes, E. *et al.* Gluten intake interferes with the humoral immune response to recombinant hepatitis B vaccine in patients with celiac disease. *Pediatrics* **121**, e1570–e1576 (2008).
82. Menendez-Corrada, R., Nettlehip, E. & Santiago-Delpin, E. A. HLA and tropical sprue. *Lancet* **2**, 1183–1185 (1986).
83. Ghoshal, U. C. *et al.* Tropical sprue is associated with contamination of small bowel with aerobic bacteria and reversible prolongation of orocecal transit time. *J. Gastroenterol. Hepatol.* **18**, 540–547 (2003).
84. Hayes, K. S. *et al.* Exploitation of the intestinal microflora by the parasitic nematode *Trichuris muris*. *Science* **328**, 1391–1394 (2010).
- This study demonstrates the co-evolution of bacterial and eukaryotic components of the microbiota and its effect on host immunity.**
85. Faith, J. J., McNulty, N. P., Rey, F. E. & Gordon, J. I. Predicting a human gut microbiota's response to diet in gnotobiotic mice. *Science* doi:10.1126/science.1206025 (19 May 2011).
86. Gaboriau-Routhiau, V., Raibaud, P., Dubuquoy, C. & Moreau, M. C. Colonization of gnotobiotic mice with human gut microflora at birth protects against *Escherichia coli* heat-labile enterotoxin-mediated abrogation of oral tolerance. *Pediatr. Res.* **54**, 739–746 (2003).
87. Mazmanian, S. K., Liu, C. H., Tzianabos, A. O. & Kasper, D. L. An immunomodulatory molecule of symbiotic bacteria directs maturation of the host immune system. *Cell* **122**, 107–118 (2005).



88. Liu, G., Yang, K., Burns, S., Shrestha, S. & Chi, H. The S1P<sub>1</sub>-mTOR axis directs the reciprocal differentiation of T<sub>H</sub>1 and T<sub>reg</sub> cells. *Nature Immunol.* **11**, 1047–1056 (2010).
89. Procaccini, C. *et al.* An oscillatory switch in mTOR kinase activity sets regulatory T cell responsiveness. *Immunity* **33**, 929–941 (2010).
90. Iwata, M. *et al.* Retinoic acid imprints gut-homing specificity on T cells. *Immunity* **21**, 527–538 (2004).
91. Siddiqui, K. R. & Powrie, F. CD103<sup>+</sup> GALT DCs promote Foxp3<sup>+</sup> regulatory T cells. *Mucosal Immunol.* **1**, S34–S38 (2008).
92. Ertesvag, A., Engedal, N., Naderi, S. & Blomhoff, H. K. Retinoic acid stimulates the cell cycle machinery in normal T cells: involvement of retinoic acid receptor-mediated IL-2 secretion. *J. Immunol.* **169**, 5555–5563 (2002).
93. Iwata, M., Eshima, Y. & Kagechika, H. Retinoic acids exert direct effects on T cells to suppress Th1 development and enhance Th2 development via retinoic acid receptors. *Int. Immunol.* **15**, 1017–1025 (2003).
94. Lemire, J. M., Adams, J. S., Sakai, R. & Jordan, S. C. 1 $\alpha$ ,25-dihydroxyvitamin D3 suppresses proliferation and immunoglobulin production by normal human peripheral blood mononuclear cells. *J. Clin. Invest.* **74**, 657–661 (1984).
95. Mora, J. R., Iwata, M. & von Andrian, U. H. Vitamin effects on the immune system: vitamins A and D take centre stage. *Nature Rev. Immunol.* **8**, 685–698 (2008).
96. Daniel, C., Sartory, N. A., Zahn, N., Radeke, H. H. & Stein, J. M. Immune modulatory treatment of trinitrobenzene sulfonic acid colitis with calcitriol is associated with a change of a T helper (Th) 1/Th17 to a Th2 and regulatory T cell profile. *J. Pharmacol. Exp. Ther.* **324**, 23–33 (2008).
97. Wang, T. T. *et al.* 1,25-Dihydroxyvitamin D<sub>3</sub> is a direct inducer of antimicrobial peptide gene expression. *J. Immunol.* **173**, 2909–2912 (2004).
98. Sigmundsdottir, H. *et al.* DCs metabolize sunlight-induced vitamin D3 to 'program' T cell attraction to the epidermal chemokine CCL27. *Nature Immunol.* **8**, 285–293 (2007).
99. Oh, D. Y. *et al.* GPR120 is an  $\omega$ -3 fatty acid receptor mediating potent anti-inflammatory and insulin-sensitizing effects. *Cell* **142**, 687–698 (2010).
100. Atarashi, K. *et al.* ATP drives lamina propria T<sub>H</sub>17 cell differentiation. *Nature* **455**, 808–812 (2008).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We are grateful to members of our laboratory, plus our colleagues C. Semenkovich and A. Shaw for many discussions. Work cited from our laboratory was supported by grants from the National Institutes of Health (DK30292, DK70977 and DK078669), the Crohn's and Colitis Foundation of America and the Bill and Melinda Gates Foundation.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at [www.nature.com/nature](http://www.nature.com/nature). Correspondence should be addressed to J.I.G. ([jgordon@wustl.edu](mailto:jgordon@wustl.edu)).

# A conditional knockout resource for the genome-wide study of mouse gene function

William C. Skarnes<sup>1</sup>, Barry Rosen<sup>1</sup>, Anthony P. West<sup>1</sup>, Manousos Koutsourakis<sup>1</sup>, Wendy Bushell<sup>1</sup>, Vivek Iyer<sup>1</sup>, Alejandro O. Mujica<sup>1†</sup>, Mark Thomas<sup>1</sup>, Jennifer Harrow<sup>1</sup>, Tony Cox<sup>1</sup>, David Jackson<sup>1</sup>, Jessica Severin<sup>1†</sup>, Patrick Biggs<sup>1†</sup>, Jun Fu<sup>2</sup>, Michael Nefedov<sup>3</sup>, Pieter J. de Jong<sup>3</sup>, A. Francis Stewart<sup>2</sup> & Allan Bradley<sup>1</sup>

Gene targeting in embryonic stem cells has become the principal technology for manipulation of the mouse genome, offering unrivalled accuracy in allele design and access to conditional mutagenesis. To bring these advantages to the wider research community, large-scale mouse knockout programmes are producing a permanent resource of targeted mutations in all protein-coding genes. Here we report the establishment of a high-throughput gene-targeting pipeline for the generation of reporter-tagged, conditional alleles. Computational allele design, 96-well modular vector construction and high-efficiency gene-targeting strategies have been combined to mutate genes on an unprecedented scale. So far, more than 12,000 vectors and 9,000 conditional targeted alleles have been produced in highly germline-competent C57BL/6N embryonic stem cells. High-throughput genome engineering highlighted by this study is broadly applicable to rat and human stem cells and provides a foundation for future genome-wide efforts aimed at deciphering the function of all genes encoded by the mammalian genome.

Following the complete sequencing of the human and mouse genomes, the functional analysis of each of the twenty thousand or so protein-coding genes remains an important goal and a major technical challenge. Several genome-wide mutagenesis strategies have been applied in the mouse, including ethyl-nitrosourea (ENU) mutagenesis, transposon mutagenesis, gene trapping and gene targeting. Gene trapping in mouse embryonic stem (ES) cells<sup>1,2</sup> has been the most productive so far, providing hundreds of thousands of random insertional mutations in more than half of the protein-coding genes in the mouse<sup>3–5</sup>. Notably, these ES cell resources can be archived indefinitely and are easily distributed to the scientific community for the purpose of generating knockout mice. However, gene-trap alleles cannot be precisely engineered and the strategy favours genes expressed in mouse ES cells.

Given the limitations of gene trapping, it is clear that the generation of a complete set of gene knockouts in the mouse will require the application of gene-targeting technology in ES cells<sup>6–8</sup>. Gene targeting can be used to engineer virtually any alteration in the mammalian genome by homologous recombination in mouse ES cells, from point mutations to large chromosomal rearrangements<sup>9,10</sup>. Over the past 20 years, gene targeting has been used to elucidate the function of more than 5,000 mammalian genes. Scaling this technology to the remainder of the genome presents numerous technical challenges and requires the production of targeted ES cells on an unprecedented scale, beyond the scope of conventional methodologies.

The first targeting pipeline for ES cells was reported several years ago before the completion of the mouse genome sequence (Velocigene)<sup>11</sup>. Bacterial artificial chromosome (BAC)-based targeting vectors were constructed to replace the coding sequence of the target gene with a *lacZ* reporter and promoter-driven selection cassette. Oligonucleotides required for the construction of targeting vectors by recombineering were based on cDNA sequences surrounding the translation initiation

and termination signals of each target gene, thus requiring no previous knowledge of the underlying genomic structure of the gene. In a single recombineering step, modified BAC clones were generated with high efficiency and used to target genes in ES cells. Correctly targeted events, which involved the deletion of up to 70-kilobases (kb) of genomic sequence, were identified using a novel high-throughput allele-counting assay. The deletion of large regions of genomic sequence, although effective for eliminating the function of the target gene, can have unintended consequences on the regulation of adjacent and distant transcriptional units<sup>12,13</sup>.

To support and accelerate progress towards the genetic analysis of all mammalian genes, large-scale knockout consortia were established in 2006 with the goal of generating a complete resource of reporter-tagged null mutations in C57BL/6 mouse ES cells<sup>14</sup>. C57BL/6 is one of the best characterized inbred strains, is the reference strain for the mouse genome sequence and breed well in the laboratory. Thus, the study of mutant alleles in a pure C57BL/6 genetic background is considered to be ideal for large-scale phenotyping efforts that will follow. Highly germline-competent ES cell lines from the C57BL/6N substrain of mice have been established for this project<sup>15–17</sup>. A common web portal providing information and access to the resource has been established<sup>18</sup>, with links to designated repositories for ordering vectors, ES cell clones and mice.

Here we describe a pipeline for the design and mass parallel construction of conditional targeting vectors by serial 96-well BAC recombineering and high-throughput gene targeting in C57BL/6 ES cells. Our pipeline is configured to create a number of useful resources en route to the generation of targeted ES cells (Supplementary Fig. 1). Ongoing large-scale production of targeted ES cell lines demonstrates rates of homologous recombination in C57BL/6 ES cells well above the historical average. Our pipeline forms the basis for the generation

<sup>1</sup>Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, UK. <sup>2</sup>Biotechnologisches Zentrum, TU Dresden, 01062 Dresden, Germany. <sup>3</sup>Children's Hospital Oakland Research Institute, Oakland, California 94609, USA. †Present addresses: Regeneron Pharmaceuticals, Inc., Tarrytown, New York, USA (A.O.M.); RIKEN Omics Science Center, Yokohama City, Japan (J.S.); Hopkirk Institute, Massey University, New Zealand (P.B.).



of thousands of *lacZ*-tagged conditional alleles for the European Conditional Mouse Mutagenesis (EUCOMM) and the National Institutes of Health Knockout Mouse (KOMP) programs as part of the international knockout effort<sup>14</sup>.

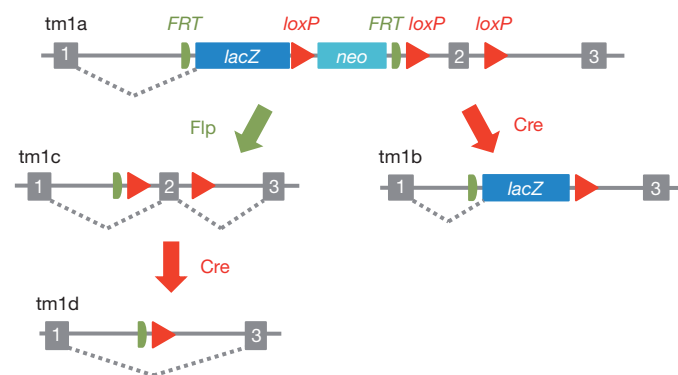
### Computer-assisted design of alleles

Conditional alleles permit the analysis of gene function in a tissue-specific or temporal manner during embryonic and postnatal development<sup>10,19</sup>. Our conditional allele is based on the 'knockout-first' design<sup>20</sup>, a strategy that combines the advantages of both a reporter-tagged and a conditional mutation (Fig. 1 and Supplementary Fig. 2). In contrast to standard conditional designs, the initial unmodified allele is predicted to generate a null allele through splicing to a *lacZ* trapping element contained in the targeting cassette. Our trapping cassettes include the mouse *En2* splice acceptor and the SV40 polyadenylation sequences, signals that have proven to be highly effective in creating null alleles in mice<sup>2,21</sup>.

The knockout-first allele can be easily modified in ES cells or in crosses to transgenic *FLP* and *cre* mice. Conditional alleles are generated by removal of the gene-trap cassette by FLP recombinase, which reverts the mutation to wild type, leaving *loxP* sites on either side of a critical exon. Subsequent exposure to Cre deletes the critical exon to induce a frameshift mutation and trigger nonsense-mediated decay of the mutant transcript. Many *cre* transgenic strains are available for the study of gene function in specific tissues and developmental time points (see <http://www.creline.org>).

Typically, *loxP* sites are placed in introns of genes to avoid disrupting normal transcription, processing and translation of the target gene. The *loxP* and *FRT* sites are positioned to minimize possible interference with the splice sites of the critical exon. In some cases, the presence of the recombinase sites may perturb normal splicing patterns<sup>22</sup>. This caveat notwithstanding, knockout first alleles are very useful for proving the causality of gene disruptions and observed phenotypes. Reversion of the phenotype with FLP, or conversely, induction of the phenotype with Cre, rule out potential effects of secondary linked mutations that can arise in cultured ES cells<sup>23</sup>. Furthermore, removal of the FRT-flanked stop cassette is particularly useful for further studies of genes that present heterozygous lethal phenotypes.

The vector design process ideally begins with high-quality manual annotation of gene structures<sup>24</sup>. Manual annotation identifies and resolves errors in automated gene predictions and captures all known transcript variants from available messenger RNA evidence. However,



**Figure 1 | Schematic of the 'knockout-first' conditional allele.** The 'knockout-first' allele (tm1a) contains an IRES-*lacZ* trapping cassette and a floxed promoter-driven *neo* cassette inserted into the intron of a gene, disrupting gene function. FLP converts the 'knockout-first' allele to a conditional allele (tm1c), restoring gene activity. Cre deletes the promoter-driven selection cassette and floxed exon of the tm1a allele to generate a *lacZ*-tagged allele (tm1b) or deletes the floxed exon of the tm1c allele to generate a frameshift mutation (tm1d), triggering nonsense mediated decay of the deleted transcript.

manual annotation of genes is a time-consuming process and proved rate-limiting in our high-throughput pipeline. Although the accuracy of automated gene prediction is improving, vector designs for Ensembl gene structures must be approached with caution.

To assist in the design of conditional alleles, we developed a computational tool to identify oligonucleotide sequences (50-mers) suitable for recombineering. These sequences are used to insert a selection cassette and *loxP* site around the critical exon and to recover homologous sequence from the BAC required for gene targeting (Fig. 2a). More generally, these computational tools can be applied to any other mammalian or non-mammalian genome for which the construction of large numbers of recombineered DNA constructs is desired. Each design is displayed on the genome browser (Fig. 2b) and manually inspected to choose the optimal design. Valid designs are selected for the 5'-most critical exon(s) that is common to all known transcript variants and disrupts at least 50% of the protein-coding sequence. Designs are rejected if the deleted region contains highly conserved intronic sequence as these elements are likely to correspond to regulatory elements and complicate the interpretation of the mutant phenotype in mice<sup>12,13</sup>.

Approximately 40% of protein-coding genes do not fit our design criteria, most commonly, small transcription units composed of one or two exons. Genes with alternative 5' end transcripts are also problematical. In some cases, it is not possible to remove a single exon or cluster of exons that disrupts all isoforms. These genes have been set aside for other partners within the international knockout consortium to generate standard *lacZ*-tagged deletion alleles using, for example, Velocigene technology<sup>11</sup>.

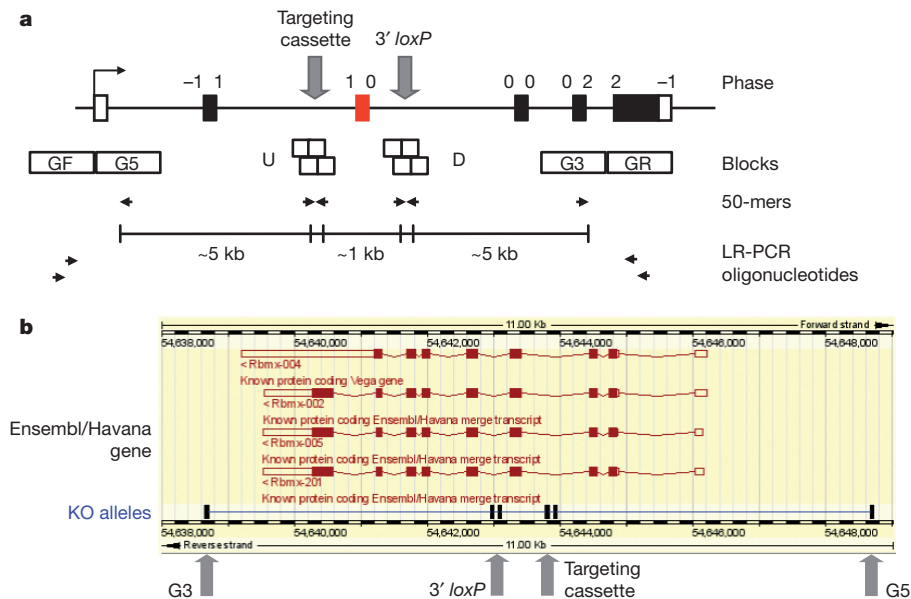
### Construction of modular targeting vectors

For the generation of conditional gene-targeting vectors, we developed a strategy for high-throughput, serial, liquid BAC recombineering in 96-well format (Fig. 3) similar to that reported for transgene production<sup>25,26</sup>. We adopted a modular strategy for the construction of targeting vectors using recombineering to create Gateway-adapted intermediate vectors (Fig. 4a) that are later assembled into the final targeting construct through *in vitro* Gateway reactions (Fig. 4b). For targeting in C57BL/6N ES cells<sup>16</sup>, we made use of indexed C57BL/6J BAC libraries<sup>27</sup> for the construction of targeting vectors.

The construction of Gateway-adapted intermediate targeting vectors from BACs involves three consecutive recombineering steps: insertion of an attR1/attR2 *zeo-pheS* Gateway element upstream of the critical exon (Fig. 3b and Supplementary Fig. 3); insertion of a floxed *kanR* cassette downstream of a critical exon (Fig. 3c); and sub-cloning of the modified region of genomic DNA (8–10 kb) into a Gateway-adapted plasmid backbone by gap repair (Fig. 3d and Supplementary Fig. 3). Heterologous attR3/attR4 sites are included to enable switching of the plasmid backbone to introduce a negative selection cassette for positive-negative targeting in ES cells. The exquisite efficiency and nucleotide precision of Red operon-induced recombination in bacteria permitted the assembly of DNA constructs in 96-well format through three rounds of recombineering with an 80% overall efficiency (Supplementary Table 1). This efficiency of vector production readily accommodates the needs of the global mouse gene-targeting projects that aim to knock out thousands of genes per year<sup>14</sup>.

### Assembly of the final targeting constructs

Gateway technology has been successfully used for the construction of large-scale genomic resources<sup>28,29</sup>. The use of Gateway technology minimizes the potential for deleterious mutations common to polymerase chain reaction (PCR)-based cloning methods. We developed a series of promoterless and promoter-driven selection cassettes flanked by attL1/attL2 sites (Supplementary Fig. 4). To use positive-negative selection for gene targeting<sup>30</sup>, a plasmid backbone was constructed that contains attL3/attL4 Gateway elements and a diphtheria-toxin-A-chain<sup>31</sup> (DTA) expression cassette. Final targeting constructs were

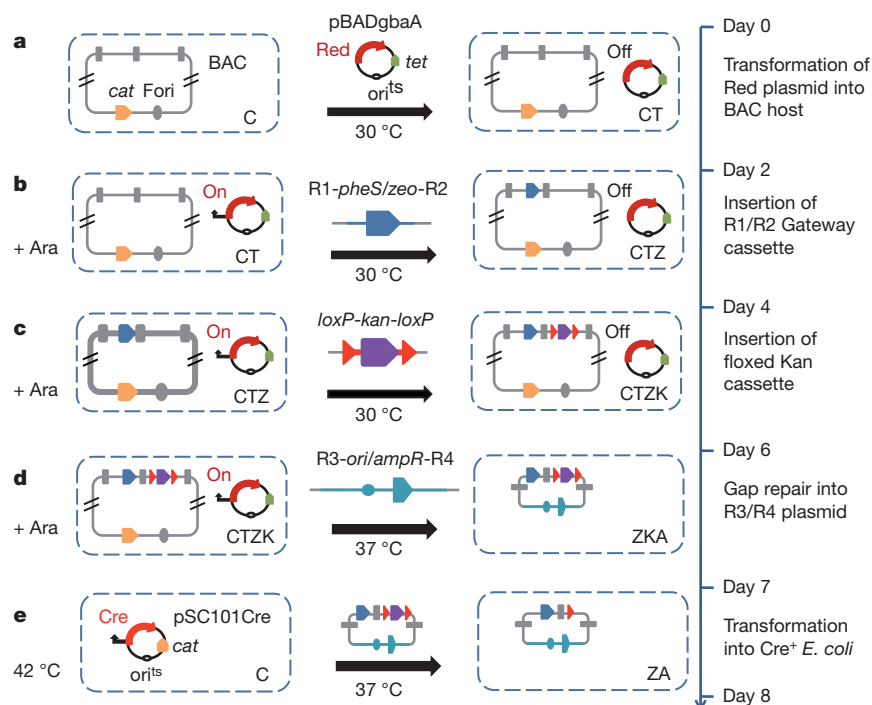


**Figure 2 | Computational design of oligonucleotides for recombineering and LR-PCR genotyping.** **a**, A critical exon(s) common to all transcript variants (red box) is identified. Recombineering oligonucleotides (50-mers) are identified by ArrayOligoSelector<sup>46</sup> within pre-defined blocks (G5, U, D, G3) of genomic sequence for insertions of the targeting cassette and 3' *loxP* site and for plasmid rescue of the 5' and 3' homology arms by gap repair. For LR-PCR

genotyping, multiple primers (25 to 30-mers) are then selected from 1-kb blocks of genomic sequence (GF, GR) outside the homology arms. **b**, Display of conditional alleles on the Ensembl genome browser (Distributed Annotation System (DAS) source = KO alleles). A conditional design for the merged Ensembl/Havana *Rbm* gene on the reverse strand is shown.

assembled *in vitro* in a three-part Gateway reaction (Fig. 4b) in 96-well format and sequence-confirmed across all recombined junctions. Final targeting vectors were recovered from 95% of the intermediate plasmids (Supplementary Table 1). Thus, the overall efficiency of vector construction is 75% and, so far, we have constructed more than 12,000 final targeting vectors.

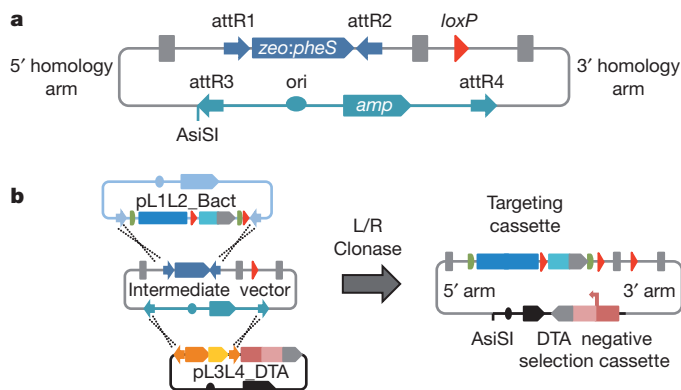
The intermediate vectors themselves (Fig. 4a) represent an important modular resource that can be re-used to generate alternative vector designs or additional mutant alleles in the future. For example, targeting cassettes containing specialized reporters, such as alkaline phosphatase or green fluorescent protein, can be rapidly assembled to provide alternative visualization of gene expression. Furthermore, targeting



**Figure 3 | Construction of Gateway-adapted intermediate targeting vectors by 96-well BAC recombineering.** Recombineering steps and elapsed time are shown. **a**, BAC clones, arrayed in 96-well format and electroporated with a plasmid expressing arabinose-inducible Red proteins (pBADgbaA)<sup>47</sup>. **b–d**, After arabinose induction, cells are electroporated with PCR fragments containing R1-*pheS/zeo*-R2 Gateway element (**b**), *loxP-kan-loxP* cassette

(**c**) and R3-ori/*ampR*-R4 subcloning plasmid (**d**). **e**, After gap repair, plasmid DNA is prepared and transformed into Cre-expressing bacteria to remove the *kanR* cassette, leaving a single *loxP* site downstream of the critical exon. Antibiotics used at each step are: A, ampicillin; C, chloramphenicol; K, kanamycin; T, tetracycline; Z, zeocin.





**Figure 4 | Intermediate and final targeting constructs.** **a**, Schematic showing the structure of the Gateway-adapted intermediate plasmid. A rare AsiSI restriction site is included in the gap repair plasmid for linearizing the final targeting vector before electroporation of ES cells. **b**, Assembly of final targeting vectors in a multi-Gateway reaction. See Supplementary Fig. 4 for a full description of the custom Gateway-adapted plasmids used for vector construction.

vectors with different selectable markers can be readily constructed to knock out the second allele of genes for functional studies in homozygous ES cells. Finally, knock-ins of wild-type and mutant cDNAs provide an avenue for detailed structure–function studies or to explore human variation. Thus, a permanent library of intermediate targeting plasmids will permit the further exploitation of targeting technology in the future.

## High-throughput ES cell production

To scale targeting experiments to high throughput, we optimized electroporation conditions for C57BL/6N ES cells<sup>16</sup> in multi-well cuvettes. Here we aimed to minimize the number of cells and amount of plasmid DNA required to obtain sufficient drug-resistant colonies for screening (Table 1). After selection, expansion and freezing, most (65%) ES cell clones retained their ability to colonize the germ line of mice<sup>16</sup>.

Homologous recombinants generated with targeting vectors are usually identified by Southern blotting. However, this method is not practical for large-scale screening. Long-range PCR (LR-PCR) is an alternative method<sup>32</sup> which is better-suited to high-throughput genotyping of ES cell clones. We developed a 384-well LR-PCR method to identify correctly targeted events (Fig. 5). PCR fragments, amplified with gene-specific primers outside the homology arms in combination with primers in the targeting cassette, were sequence-verified. In general, LR-PCR was performed across the 3' homology arm. Because the targeted clones are genotyped at one end, non-homologous events within the opposite arm will occur in rare cases. Furthermore, mixed clones composed of targeted and non-targeted cells are not detected by our high-throughput genotyping protocol. For these reasons, further validation of targeted alleles using standard Southern blot assays is highly recommended before use.

Owing to frequent crossover events between the selectable marker and 3' *loxP* site, many of the targeted ES cell clones lose the 3' *loxP* site and cannot be converted to a conditional allele. To distinguish between these two alternative products of homologous recombination, LR-PCR products amplified from the 3' homology arm were sequenced with a primer at the *loxP* site. Where 3' LR-PCR failed

to generate a product, LR-PCR was performed across the 5' homology arm (5' LR-PCR). For these cases, the retention of the 3' *loxP* site was confirmed by PCR between the cassette and 3' *loxP* site.

## Gene targeting is highly efficient

High-throughput gene targeting depends on achieving high targeting efficiencies. For genes expressed in ES cells, a promoterless targeting strategy (referred to as 'targeted trapping')<sup>33</sup> has been shown to yield targeting efficiencies averaging above 50%. By design, promoterless vectors effectively suppress the recovery of random non-homologous events in the genome as only insertions in transcribed loci, in the correct orientation and reading frame, will confer drug resistance. We electroporated 1,285 different promoterless constructs and obtained targeted clones from nearly half of these constructs with an average targeting efficiency of 50% (Table 1). These data confirm and extend the results of ref. 33, demonstrating that targeted trapping is a highly efficient method for genes expressed in ES cells.

Only half of the promoterless targeting vectors were effective in producing targeted clones. Electroporation of these vectors produced variable numbers of drug-resistant colonies. In general, high colony numbers were predictive of successful targeting experiments, whereas low colony numbers usually indicated a failure to target the locus (Supplementary Table 2). The success or failure of a construct correlated with the number of clones with gene-trap events in the International Gene Trap Consortium database (Supplementary Table 3). Thus, gene-trapping data serve as a useful guide to identify the subset of genes that are amenable to a promoterless targeting strategy<sup>34</sup>. Correlation with classes of gene was also observed. For instance, targeted trapping was less effective with secreted proteins compared to non-secreted proteins, indicating that our cassette designed for trapping secreted proteins (pL1L2\_ST, see Supplementary Fig. 4)<sup>35</sup> is not optimal for this class of gene<sup>36</sup>.

Given that only half of all genes are expressed at a sufficient level in ES cells to support a targeted trapping strategy, we switched to using a promoter-driven cassette for positive selection for non-expressed genes combined with negative DTA selection to select against random insertions. We electroporated different positive–negative targeting cassettes and from the analysis of approximately 30 ES cell clones per unique construct, we recovered targeted events for 80% of genes with an average targeting efficiency of 35% (Table 1; for a complete list of targeted genes see Supplementary Data). A combination of factors probably contribute to our high targeting efficiencies, including the use of isogenic DNA, relatively long recombiner homology arms and DTA negative selection.

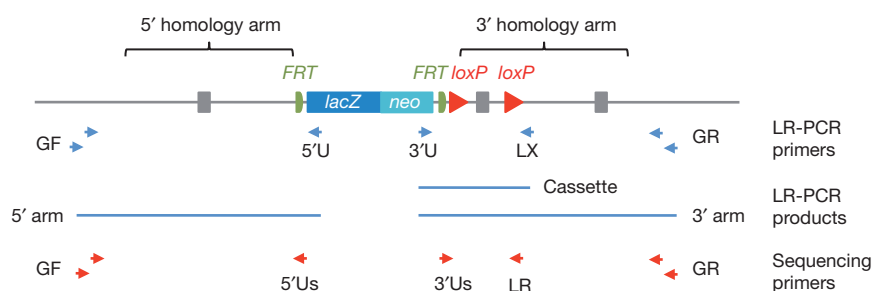
Gene targeting is dependent on both the length and the extent of homology between the targeting vector and the target locus<sup>37–39</sup>. Our vectors typically contain 10 kb of homology to the endogenous locus and originate from a C57BL/6J BAC library. Although the ES cells are derived from the C57BL/6N sub-strain, the Jackson (J) and NIH (N) substrains of C57BL/6 are very closely related<sup>16</sup>, thus our targeting vectors will have identical sequence with the ES cell genome in the great majority of cases. Negative selection was introduced to improve targeting efficiencies<sup>30,31</sup>. Overall we observed a threefold enrichment of targeted clones with DTA counter-selection, consistent with previous observations<sup>30,31,40</sup> (Table 1).

In a high-throughput pipeline, projects inevitably fail at one or more steps and overall pipeline efficiency depends on effective recovery of

**Table 1 | Targeting efficiency using promoterless and promoter-driven cassettes**

Vector type	Number of unique targeting vectors	Number of successful electroporations	Number of colonies*	Number of genes targeted	Genes targeted (%)	Targeting efficiency (%)	Number of colonies screened*	Number of targeted clones*	Number of targeted clones with 3' <i>loxP</i> site*
Promoterless	1,285	778	224	621	48	51	24	12	6
Promoter	1,811	1,671	348	1,440	80	35	29	10	3.5
Promoter (–DTA)	87	87	729	49	56	12	34	4	1

\* Average values.



**Figure 5 | Genotyping ES clones by LR-PCR sequencing.** Five LR-PCR reactions are carried out: two 5' arm (GF/5'U), two 3' arm (3'U/GR) and one cassette (3'U/LX). Sequence verification of LR-PCR products is carried out with gene-specific primers (GF and GR) and with nested primers in the

targeting cassette (5'Us and 3'Us). To confirm the presence or absence of the 3' *loxP* site, 3' arm LR-PCR products are sequenced with a primer adjacent to the *loxP* site (LR). In cases where 3' arm LR-PCR fails to generate a product, the 3' *loxP* site is confirmed by sequencing the cassette product.

these failures. In our experience, most failures are technical in nature and are most efficiently recovered by repeating the procedure. For example, 70% of targeting experiments are rescued after re-electroporation of cells with an alternative preparation of vector DNA (Supplementary Data). Similarly, re-synthesis of oligonucleotides for recombineering or repeating the Gateway reaction recovers a majority of intermediate and final targeting vectors (data not shown). Thus, completion of the mutant resource will require iterative rounds of recovery. Whether some genes are refractory to targeting will become apparent once all technical issues have been ruled out.

## Discussion

Our targeting pipeline is the major contributor to the international mouse knockout programmes that aim to generate *lacZ*-tagged null mutations in every protein-coding gene in mouse. With the technology described here, more than 9,000 genes have been successfully targeted in C57BL/6N ES cells to date. The value of our knockout ES cell resource critically depends on the germline potential of individual targeted C57BL/6N ES cell clones. In a separate study<sup>16</sup>, hundreds of targeted cell lines generated in our pipeline were assessed for contribution to the germline after blastocyst injection. At least 65% of targeted clones colonized the germ line of chimaeric mice. Thus, our library of mutant C57BL/6N ES cells is robust and will support the production of mutant mice for future large-scale phenotyping programmes.

The scale of mass parallel vector construction and gene targeting described here has implications for functional genomics and proteomics in many model systems. New systematic, genome-scale programmes can now be contemplated. Using available BAC or fosmid genome resources, the high-throughput production of complex transgenes and/or targeting constructs will facilitate the generation of sophisticated, physiologically accurate, cell and animal models. For example, tagging all proteins in the mouse genome by knock-in targeting to establish a proteomic mapping programme equivalent to the highly successful yeast TAP-tagging programmes<sup>41</sup> is now feasible.

In the coming years, it is likely that the genome engineering technologies pioneered in the mouse will be also applicable to other model systems such as the rat<sup>42,43</sup> and human pluripotent stem cells<sup>44,45</sup>. The capacity for fluent gene targeting also permits the systematic generation of doubly targeted ES cell lines for functional studies by conditional mutagenesis, which will serve to complement and extend RNA interference studies by providing complete genetic knockouts. Coupled with the power to differentiate ES cells into many cell types, such resources will not only provide means to gaining unique functional insights but will also reduce animal experimentation. With pioneering methodologies, we have overcome the considerable technical challenges involved in establishing the most complex and accurate high-throughput functional genomics platform yet attempted. We believe that our work raises the standards of achievement and expectation for future genome-scale programmes.

## METHODS SUMMARY

**Gene annotation and vector design software.** Manual annotation of mouse gene structures was carried out as previously described<sup>24</sup>. Vector designs are based on the current release of the Ensembl and Vega databases (NCBIM37 assembly). Critical exon(s) for each target gene are identified computationally (start phase – end phase = 0). Using ArrayOligoSelector<sup>46</sup>, our software returns a set of six 50-mer oligonucleotides at defined distances from the critical exon(s) for recombineering. **96-well recombineering and three-way Gateway reactions.** BAC clones are ordered from indexed C57BL/6J libraries<sup>27</sup> (RP23/24), arrayed in 96-well plates and transformed with pBADgbaA<sup>47</sup> plasmid encoding lambda Red recombination proteins. Three rounds of recombineering are carried out serially in 96-well cultures using DNA cassettes amplified by PCR with primers containing 50-nucleotide homology to target sequences<sup>25,26</sup>. After gap repair, plasmid DNA is transformed into Cre-expressing bacteria to reduce the floxed *kanR* cassette to a single *loxP* site.

Three-way Gateway reactions containing intermediate vector, attL1/attL2 targeting cassette and attL3/attL4 DTA plasmids are incubated with LR Clonase II Plus (Invitrogen), transformed into bacteria and selected on agar plates containing appropriate antibiotics and 4-chlorophenylalanine<sup>48</sup>. Final targeting constructs are sequence-verified across each recombineered junction, linearized with *Asi*SI and visualized on E-Gels (Invitrogen) to verify their size (Supplementary Fig. 5).

**Electroporation of ES cells and LR-PCR genotyping.** Electroporation of C57BL/6N mouse ES cells<sup>16</sup> with linearized plasmid DNA was carried out in 25-well electroporation cuvettes (BTX Harvard Apparatus). Stable clones were selected in medium containing Geneticin (Invitrogen). Typically 32 clones are picked, expanded in 96-well plates and archived in 96-well cryovials (Matrix).

Long-range PCR reactions using SequelPrep (Life Technologies) or LongAMP (NEB) were carried out with genomic DNA from direct lysis of ES cells grown in 96-well plates. PCR products were visualized on E-gels (Supplementary Fig. 6) then treated with exonuclease I and phosphatase (NEB) and sequenced.

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Received 3 August 2009; accepted 27 April 2011.**

- Gossler, A., Joyner, A. L., Rossant, J. & Skarnes, W. C. Mouse embryonic stem cells and reporter constructs to detect developmentally regulated genes. *Science* **244**, 463–465 (1989).
- Skarnes, W. C., Auerbach, B. A. & Joyner, A. L. A gene trap approach in mouse embryonic stem cells: the *lacZ* reported is activated by splicing, reflects endogenous gene expression, and is mutagenic in mice. *Genes Dev.* **6**, 903–918 (1992).
- Zambrowicz B. P. *et al.* Wnk1 kinase deficiency lowers blood pressure in mice: a gene-trap screen to identify potential targets for therapeutic intervention. *Proc. Natl Acad. Sci. USA* **100**, 14109–14114 (2003).
- International Gene Trap Consortium. A public gene trap resource for mouse functional genomics. *Nature Genet.* **36**, 543–544 (2004).
- Hansen, G. M. *et al.* Large-scale gene trapping in C57BL/6N mouse embryonic stem cells. *Genome Res.* **18**, 1670–1679 (2008).
- Bradley, A., Evans, M., Kaufman, M. H. & Robertson, E. Formation of germ-line chimaeras from embryo-derived teratocarcinoma cell lines. *Nature* **309**, 255–256 (1984).
- Robertson, E., Bradley, A., Kuehn, M. & Evans, M. Germ-line transmission of genes introduced into cultured pluripotent cells by retroviral vector. *Nature* **323**, 445–448 (1986).
- Thomas, K. R. & Capecchi, M. R. Site-directed mutagenesis by gene targeting in mouse embryo-derived stem cells. *Cell* **51**, 503–512 (1987).

9. van der Weyden, L., Adams, D. J. & Bradley, A. Tools for targeted manipulation of the mouse genome. *Physiol. Genomics* **11**, 133–164 (2002).
10. Glaser, S., Anastasiadis, K. & Stewart, A. F. Current issues in mouse genome engineering. *Nature Genet.* **37**, 1187–1193 (2005).
11. Valenzuela, D. M. *et al.* High-throughput engineering of the mouse genome coupled with high-resolution expression analysis. *Nature Biotechnol.* **21**, 652–659 (2003).
12. Lettice, L. A. *et al.* Disruption of a long-range cis-acting regulator for Shh causes preaxial polydactyly. *Proc. Natl Acad. Sci. USA* **99**, 7548–7553 (2002).
13. Zuniga, A. *et al.* Mouse limb deformity mutations disrupt a global control region within the large regulatory landscape required for Gremlin expression. *Genes Dev.* **18**, 1553–1564 (2004).
14. International Mouse Knockout Consortium. A mouse for all reasons. *Cell* **128**, 9–13 (2007).
15. Poueymirou, W. T. *et al.* F0 generation mice fully derived from gene-targeted embryonic stem cells allowing immediate phenotypic analyses. *Nature Biotechnol.* **25**, 91–99 (2007).
16. Pettitt, S. J. *et al.* Agouti C57BL/6N embryonic stem cells for mouse genetics resources. *Nature Methods* **6**, 493–495 (2009).
17. Gertsenstein, M. *et al.* Efficient generation of germ line transmitting chimeras from C57BL/6N ES cells by aggregation with outbred host embryos. *PLoS ONE* **5**, e11260 (2010).
18. Ringwald, M. *et al.* The IKMC web portal: a central point of entry to data and resources from the International Knockout Mouse Consortium. *Nucleic Acids Res.* **39** (Database issue), D849–D855 (2011).
19. Branda, C. S. & Dymecki, S. M. Talking about a revolution: The impact of site-specific recombinases on genetic analyses in mice. *Dev. Cell* **6**, 7–28 (2004).
20. Testa, G. *et al.* A reliable lacZ expression reporter cassette for multipurpose, knockout-first alleles. *Genesis* **38**, 151–158 (2004).
21. Mitchell, K. J. *et al.* Functional analysis of secreted and transmembrane proteins critical to mouse development. *Nature Genet.* **28**, 241–249 (2001).
22. Yang, S. H. *et al.* Caution! Analyze transcripts from conditional knockout alleles. *Transgenic Res.* **18**, 483–489 (2009).
23. Liang, Q., Conte, N., Skarnes, W. C. & Bradley, A. Extensive genomic copy number variation in embryonic stem cells. *Proc. Natl Acad. Sci. USA* **105**, 17453–17456 (2008).
24. Wilming, L. G. *et al.* The vertebrate genome annotation (Vega) database. *Nucleic Acids Res.* **36**, D753–D760 (2008).
25. Sarov, M. *et al.* A recombineering pipeline for functional genomics applied to *Caenorhabditis elegans*. *Nature Methods* **3**, 839–844 (2006).
26. Poser, I. *et al.* BAC TransGenomics: a high-throughput method for exploration of protein function in mammals. *Nature Methods* **5**, 409–415 (2008).
27. Osoegawa, K. *et al.* Bacterial artificial chromosome libraries for mouse sequencing and functional analysis. *Genome Res.* **10**, 116–128 (2000).
28. Hope, I. A. *et al.* Feasibility of genome-scale construction of promoter:reporter gene fusions for expression in *Caenorhabditis elegans* using a multisite gateway recombination system. *Genome Res.* **14**, 2070–2075 (2004).
29. Rual, J. F. *et al.* Towards a proteome-scale map of the human protein–protein interaction network. *Nature* **437**, 1173–1178 (2005).
30. Mansour, S. L., Thomas, K. R. & Capecchi, M. R. Disruption of the proto-oncogene *int-2* in mouse embryo-derived stem cells: a general strategy for targeting mutations to non-selectable genes. *Nature* **336**, 348–352 (1988).
31. Yagi, T. *et al.* Homologous recombination at c-fyn locus of mouse embryonic stem cells with use of diphtheria toxin A-fragment gene in negative selection. *Proc. Natl Acad. Sci. USA* **87**, 9918–9922 (1990).
32. Lay, J. M., Friis-Hansen, L., Gillespie, P. J. & Samuelson, L. C. Rapid confirmation of gene targeting in embryonic stem cells using two long-range PCR techniques. *Transgenic Res.* **7**, 135–140 (1998).
33. Friedel, R. H. *et al.* Gene targeting using a promoterless gene trap vector (“targeted trapping”) is an efficient method to mutate a large fraction of genes. *Proc. Natl Acad. Sci. USA* **102**, 13188–13193 (2005).
34. Skarnes, W. C. Two ways to trap a gene in mice. *Proc. Natl Acad. Sci. USA* **102**, 13001–13002 (2005).
35. Skarnes, W. C., Moss, J. E., Hurtley, S. M. & Beddington, R. S. Capturing genes encoding membrane and secreted proteins important for mouse development. *Proc. Natl Acad. Sci. USA* **92**, 6592–6596 (1995).
36. de Felipe, P., Luke, G. A., Brown, J. D. & Ryan, M. D. Inhibition of 2A-mediated ‘cleavage’ of certain artificial polyproteins bearing N-terminal signal sequences. *Biotechnol. J.* **5**, 213–223 (2010).
37. Hastay, P., Rivera-Perez, J. & Bradley, A. The length of homology required for gene targeting in embryonic stem cells. *Mol. Cell. Biol.* **11**, 5586–5591 (1991).
38. Deng, C. & Capecchi, M. R. Reexamination of gene targeting frequency as a function of the extent of homology between the targeting vector and the target locus. *Mol. Cell. Biol.* **12**, 3365–3371 (1992).
39. te Riele, H., Maanday, E. R. & Berns, A. Highly efficient gene targeting in embryonic stem cells through homologous recombination with isogenic DNA constructs. *Proc. Natl Acad. Sci. USA* **89**, 5128–5132 (1992).
40. Yanagawa, Y. *et al.* Enrichment and efficient screening of ES cells containing a targeted mutation: the use of DT-A gene with the polyadenylation signal as a negative selection marker. *Transgenic Res.* **8**, 215–221 (1999).
41. Collins, S. R. *et al.* Toward a comprehensive atlas of the physical interactome of *Saccharomyces cerevisiae*. *Mol. Cell. Proteomics* **6**, 439–450 (2007).
42. Buehr, M. *et al.* Capture of authentic embryonic stem cells from rat blastocysts. *Cell* **135**, 1287–1298 (2008).
43. Li, P. *et al.* Germline competent embryonic stem cells derived from rat blastocysts. *Cell* **135**, 1299–1310 (2008).
44. Thomson, J. A. *et al.* Embryonic stem cell lines derived from human blastocysts. *Science* **282**, 1145–1147 (1998).
45. Takahashi, K. *et al.* Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell* **131**, 861–872 (2007).
46. Bozdech, Z. *et al.* Expression profiling of the schizont and trophozoite stages of *Plasmodium falciparum* with a long-oligonucleotide microarray. *Genome Biol.* **4**, R9 (2003).
47. Wang, J. *et al.* An improved recombineering approach by adding RecA to lambda Red recombination. *Mol. Biotechnol.* **32**, 43–53 (2006).
48. Kast, P. pKSS—a second-generation general purpose cloning vector for efficient positive selection of recombinant clones. *Gene* **138**, 109–114 (1994).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank the following people for technical assistance: D. Klose, D. Oakley, W. Yang and L. Stebbings for informatics/vector design; R. Bennett, A. Horton and A. van Brunt for manual gene annotation/vector design; L. Cho, R. Li, J.-F. Popoff, M. Sharma and Y. Zhang for recombineering; G. Belteki, P. Tate, Y. Bekele and S. Borchia for targeting vectors; D. Fraser, J. Greystrom, N. Gueorguieva, M. Jackson, P. Ramagiri, I. Walczak, J. Woodward, E. Stebbings, M. Martinez, A. Tsang and Y. Yoshinaga for vector/ES quality control; and D. Edwards, S. Harris, N. Krishnappa, R. Leah and A. Tait for ES cells. We are grateful for advice on the Gateway system from J. Chesnut of Invitrogen. Finally, we wish to thank W. Wurst, K. Lloyd, and our EUComm and KOMP colleagues who are contributing to the production and distribution of the conditional knockout resource. This work was funded by the Wellcome Trust Sanger Institute, grants from the National Institutes of Health (KOMP, U01-HG004080 to W.C.S., P.J.d.J. and A.B.) from the EU Sixth Framework Programme (EUComm, to W.C.S., A.F.S. and A.B.).

**Author Contributions** W.C.S., B.R., A.P.W., M.K., W.B. and A.O.M. designed the experiments and contributed equally to this work. V.I. and T.C. developed the vector design software. A.O.M., M.T. and J.H. performed and managed manual curation of gene structures and selection of conditional designs. The modular design of targeting vectors was conceived by B.R. Recombineering of vectors was developed by B.R., W.C.S., M.K., M.N. and P.J.d.J., and managed by M.K. and P.J.d.J. Recombineering reagents and advice were supplied by J.F. and A.F.S. High-throughput targeting of ES cells was developed by W.C.S. and managed by W.B. Sequence confirmation of vectors and genotyping of targeted ES cell clones was developed and managed by A.P.W., with informatic support from V.I., D.J., J.S. and P.B. A.B. and A.F.S. inspired the work and wrote the paper together with W.C.S. All authors read and provided comments on the final manuscript.

**Author Information** Detailed information on targeted genes is available from the IKMC web portal (<http://www.knockoutmouse.org>). Targeting constructs and mutant ES cells are available upon request from the EUComm (<http://www.eummc.org>) and KOMP (<http://www.komp.org>) repositories. Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at [www.nature.com/nature](http://www.nature.com/nature). Correspondence and requests for materials should be addressed to W.C.S. ([skarnes@sanger.ac.uk](mailto:skarnes@sanger.ac.uk)).



## METHODS

**Computational design of conditional alleles.** Gene structures to be targeted are first extracted from a current release of the Ensembl (NCBIM37 assembly) or Vega database. Critical exons, which when deleted induce a frameshift, are chosen computationally (start phase – end phase  $\neq 0$ ) or manually (exon length not divisible by 3). Primers (50-mer oligonucleotides) for recombineering are then selected from overlapping blocks of sequence (typically 120 bp) flanking the critical exons at a predefined distance from the splice sites (300 bp from the splice acceptor and 100 bp from the splice donor). Primers for gap repair were chosen from sequence blocks (typically 1 kb) at the ends of the desired homology arms (4–6 kb). Each block was analysed by ArrayOligoSelector<sup>46</sup> (<http://sourceforge.net/projects/arrayoligosel/>) generating one or more candidate primers inside each sequence block with a minimum of 28% G+C content. Candidate primers were rejected if they were repetitive inside a region spanning 100 kb either side of the critical exon(s), and gap repair primers at the ends of the homology arms were also rejected if they shared sequences of 6 bp or more. The final recombineering primer sequences were mapped to the current NCBI assembly, recorded with their genomic coordinates in a database, and displayed in an Ensembl DAS-track. After manual inspection, complete sets of recombineering primers were selected from the database, automatically reverse complemented (where appropriate) and appended with 20–23 bp of sequence homology to the appropriate recombineering cassettes before ordering. In parallel, BACs from the RP23/RP24 indexed library were chosen based on end-mappings of the clones. A vector design interface (Custom Design Tool; <http://www.sanger.ac.uk/htgt/>) is available online.

**96-well recombineering.** BACs from the RPCI-23/RPCI-24 indexed C57BL/6J libraries<sup>27</sup> were arrayed in 96-well format to match the corresponding 96-well plates of 70-mer oligonucleotides (desalted; Illumina/Invitrogen) used to PCR amplify the cassettes used for recombineering. PCR amplifications were performed using the FastStart High Fidelity PCR System (Roche) and the products were desalted using High Pure 96UF Cleanup kits (Roche). The arrayed BAC clones were initially grown at 37 °C in Luria broth (LB) containing chloramphenicol (12.5  $\mu\text{g ml}^{-1}$ ) to early log phase and made electrocompetent by washing three times with ice-cold HPLC grade water and the cells are transformed with pBADgbaA plasmid DNA<sup>47</sup> using an ECM 630 96-well electroporator/HT-200 automatic plate handler (BTX Harvard Apparatus; pulse conditions of 2,400 V, 700  $\Omega$ , 25  $\mu\text{F}$ ) followed by growth at 30 °C in liquid medium containing tetracycline (5  $\mu\text{g ml}^{-1}$ ) and chloramphenicol (12.5  $\mu\text{g ml}^{-1}$ ). The BAC cultures underwent three rounds of recombineering, changing only the PCR products used for each electroporation and the antibiotic selection applied after each step, using the following standard procedure: early log phase cultures were induced to express the red operon following addition of 0.1% arabinose and incubated for 40 min at 37 °C; electrocompetent cells were electroporated in 96-well format (as above) with 1–2  $\mu\text{g}$  of desalted PCR products and allowed to recover at 37 °C for 90 min; an aliquot was then inoculated into a new 96-well box containing media plus the appropriate antibiotics and grown at 30 °C for 2 days. The PCR cassette and antibiotic cocktail used at each step shown in Fig. 3 was as follows. (1) R1-*pheS/zeo-R2*, zeocin (4  $\mu\text{g ml}^{-1}$ ), tetracycline (5  $\mu\text{g ml}^{-1}$ ), chloramphenicol (12.5  $\mu\text{g ml}^{-1}$ ); (2) *loxP-kan-loxP*, kanamycin (15  $\mu\text{g ml}^{-1}$ ), zeocin (6.5  $\mu\text{g ml}^{-1}$ ), tetracycline (5  $\mu\text{g ml}^{-1}$ ), chloramphenicol (12.5  $\mu\text{g ml}^{-1}$ ); and (3) pR3R4, zeocin (6.5  $\mu\text{g ml}^{-1}$ ), kanamycin (15  $\mu\text{g ml}^{-1}$ ) carbenicillin (50  $\mu\text{g ml}^{-1}$ ). After the gap repair step, the temperature was shifted to 37 °C to eliminate the recombineering plasmid. Intermediate plasmid DNA was purified using standard procedures from saturated cultures (1.5 ml) grown in 96-well blocks. Approximately 50 ng was transformed into electrocompetent DH10B *E. coli* carrying the 705-Cre plasmid (Gene Bridges), pre-induced at 42 °C to express Cre recombinase from the  $\lambda_{\text{PR}}$  promoter, and selected in liquid culture containing carbenicillin (50  $\mu\text{g ml}^{-1}$ ) and zeocin (10  $\mu\text{g ml}^{-1}$ ). After overnight growth at 37 °C, individual colonies were streaked out on ampicillin/zeocin plates to isolate individual clones and were sequence-verified.

**Gateway exchange reaction.** Three-way Gateway reactions were carried out in 96-well format using LR Clonase II Plus enzyme mix (Invitrogen) essentially as described by the manufacturer. In an overnight reaction at 25 °C, 100–200 ng of intermediate targeting vector (prepared from 1.5-ml cultures in 96-well blocks using the Qiagen Turboprep kit) was combined with 60 ng of L1/L2 targeting cassette vector and 60 ng of L3/L4 DTA plasmid backbone in a 10  $\mu\text{l}$  volume. After treatment with Proteinase K, 2  $\mu\text{l}$  of the reaction was transformed into 30  $\mu\text{l}$  of chemically competent *Escherichia coli* (DH10B, Invitrogen) and plated onto YEG agar plates containing 4-chlorophenylalanine<sup>48</sup> and the appropriate antibiotics. Individual colonies were picked and sequenced across all recombinered junctions. Reads were automatically aligned against the synthetic vector sequences and assigned pass levels based on the number and position of matching reads.

**ES cell culture and electroporation.** The final targeting constructs were prepared for ES cell electroporation from 2 ml of culture (2X LB plus antibiotics) in 96-well format using the Qiagen Turboprep kit. Before electroporation, vectors were

linearized with AsiSI and examined by gel electrophoresis. For most clones, the digested DNA migrated as a single high-molecular-mass band of the expected size (Supplementary Fig. 5). Occasionally, contaminating smaller molecular mass bands were also observed on the gel (DNA quality failures).

JM8 mouse ES cell lines derived from the C57BL/6N strain were grown either on a feeder layer of SNL6/7 fibroblasts (neomycin and/or puromycin resistant) or on gelatinized tissue culture plates<sup>16</sup>. Both feeder-independent and feeder-dependent lines were maintained in Knockout DMEM (500 ml, Gibco) supplemented with 2 mM glutamine, 5 ml 100 $\times$   $\beta$ -mercaptoethanol (360  $\mu\text{l}$  in 500 ml PBS, filter sterilized), 10–15% fetal calf serum respectively (Invitrogen) and 500 U  $\text{ml}^{-1}$  leukaemia-inhibitory factor (ESGRO, Millipore). Trypsin solution was prepared by adding 20 ml of 2.5% trypsin solution (Gibco) and 5 ml chicken serum (Gibco) to 500 ml filter-sterilized PBS containing 0.1 g EDTA (Sigma) and 0.5 g D-glucose (Sigma).

Electroporations of ES cells were carried out in a 25-well cuvette using the ECM 630 96-well electroporator /HT-200 automatic plate handler (BTX Harvard Apparatus; set at 700 V, 400  $\Omega$ , 25  $\mu\text{F}$ ). Immediately before electroporation, cell suspensions of  $\sim 1 \times 10^7$  cells and  $\sim 2 \mu\text{g}$  of linearized targeting vector DNA were mixed in a final volume of 120  $\mu\text{l}$  PBS. Cells were seeded onto a 10-cm dish (with feeders or gelatin) and colonies were picked after 10 d of selection in 100  $\mu\text{g}$  (active) per ml Geneticin (Invitrogen). To expand cells into duplicate wells for archiving and preparation of genomic DNA, confluent cultures of JM8 ES cells grown on feeder cells were washed twice with pre-warmed PBS and trypsinized for 15 min at 37 °C. Five volumes of pre-warmed media were added and the cells were gently dispersed by titration and passed at a dilution of 1:4 into new plates containing feeder cells. Passage of cells grown on gelatinized plates was carried out in a similar manner except that the cells were trypsinized for 10 min and passed at a dilution of 1:6 into freshly gelatin-coated plates (0.1% gelatin, Sigma G1393). Culture medium was replaced daily and cells reached confluence 2 days after passage. To archive ES cell clones, trypsinized cells from confluent 96-well plates were transferred in 200  $\mu\text{l}$  freezing medium (Knockout DMEM, 15% serum/10% DMSO) to 96-well cryovials (Matrix) and overlaid with sterile mineral oil. The cells were placed at  $-80$  °C overnight and then transferred to liquid nitrogen.

**Computational design of primers for long-range PCR.** To identify targeted ES cell clones, we developed a robust LR-PCR system that uses one set of reaction conditions for every targeted allele screened. In addition, we used an in-house primer generation program (“Primer Brain”) to generate genome-specific primers for the LR-PCR. Primers were selected from 2-kb blocks of sequence upstream of the 5' homology arm (GF) and downstream of the 3' homology arm (GR) and from a variable-sized region that contains the critical exon (EX). Primers were first extracted by a single-base-pair tiling of each region into 24- to 30-mers that end in G/C, have at least 10 G/C bases and have a melting temperature of at least 64 °C. Primer choice was weighted negatively to avoid both ‘runs’ of nucleotides (for example, ‘AAA’) and self-annealing ends. The top 100 high-scoring primers in each region were aligned against the current mouse genome (NCBIM37) with Exonerate software (<http://www.ebi.ac.uk/~guy/exonerate>) and were weighted negatively based on the number of alignments to the genome, with added negative weight given to alignments close to the 3' end of primers. The two best-scoring primers from each block (GF1 and GF2; GR1 and GR2; EX1 and EX2) were grouped and primer combinations (for example, GF1 and EX1) were screened to eliminate pairs with a 4-bp overlap at their 3' ends. The resulting GF, GR and EX primers were stored in an Oracle database.

**LR-PCR genotyping.** ES cell genomic DNA was isolated by digesting the cells with Proteinase K and RNase A. Each well of a confluent 96-well plate was lysed with 30  $\mu\text{l}$  of lysis buffer (10 mM Tris/HCl pH 8, 1 mM EDTA, 50 mM KCl, 2 mM  $\text{MgCl}_2$ ) containing 200  $\mu\text{g ml}^{-1}$  RNase A (Sigma) and 0.67  $\text{mg ml}^{-1}$  proteinase K (Life Technologies). After overnight digestion at 60 °C, the samples were heated to 90 °C (2 min) and 1–2  $\mu\text{l}$  of the lysate was used in a 10  $\mu\text{l}$  LR-PCR reaction. To generate LR-PCR amplicons, two genomic-specific primers outside each end of the 5' and 3' homology arms (GF and GR, respectively) were used in combination with the appropriate universal cassette primers (5U (5'-CACACGGGTTC TTCTGTTAGTCC-3') and 3U (5'-ATCCGGGGTACCGCGTCGAG-3')) (Fig. 5).

Using the SequalPrep kit (0.1  $\mu\text{l}$  100% v/v DMSO, 0.5  $\mu\text{l}$  10 $\times$  enhancer A, 0.5  $\mu\text{l}$  10 $\times$  enhancer B, 1.0  $\mu\text{l}$  10 $\times$  buffer, 0.2  $\mu\text{l}$  Taq Enzyme/dNTPs; Life Technologies) or LongAMP Taq mix (0.2  $\mu\text{l}$  100% v/v DMSO (Sigma), 0.3  $\mu\text{l}$  10 mM dNTPs (Thermo Fisher Scientific), 2.0  $\mu\text{l}$  5 $\times$  LongAMP buffer (NEB), 0.4  $\mu\text{l}$  LongAMP Taq (NEB)), 10  $\mu\text{l}$  reactions were set up in 384-well format with  $\sim 30$ –50 ng (1–2  $\mu\text{l}$ ) genomic DNA and 12 pmol of each primer. Thermal cycling was performed using the following conditions: 1 cycle 93 °C for 3 min; 8 cycles 92 °C for 15 s, 65 °C for 30 s decreasing by 1 °C per cycle, 65 °C (LongAMP) or 68 °C (SequalPrep) for 8 min; 30 cycles 92 °C for 15 s, 55 °C for 30 s, 65 °C (LongAMP) or 68 °C (SequalPrep) for 8 min increasing 20 s per cycle; 1 cycle 65 °C (LongAMP) or

68 °C (SequalPrep) for 9 min. The PCR products were visualized on 1% E-gels (Life Technologies) and scored for the presence of high-molecular-mass fragments (Supplementary Fig. 6). The LR-PCR products were treated with exonuclease I and shrimp alkaline phosphatase (0.3 U  $\mu\text{l}^{-1}$  and 0.19 U  $\mu\text{l}^{-1}$ , respectively; NEB) in 20 mM Tris/HCl, 10 mM  $\text{MgCl}_2$  for 1 h at 37 °C followed by 80 °C for 15 min. PCR products were sequenced with the genomic primers used for amplification and universal primers to the targeting cassette (5'Us (5'-CGTGGTATCGT

TATGCGCCT-3') and 3'Us (5'-TCTATAGTCGCAGTAGGCGG-3')) and 3' *loxP* (LR (5'-ACTGATGGCGAGCTCAGACC-3')). Sequence reads were compared by BLAST against synthetic sequences for each targeted allele and clones with correctly aligned sequences were marked as valid. Clones that retained the 3' *loxP* site and have 3' or 5' sequence-verified LR-PCR bands are marked for distribution and clones that have lost the 3' *loxP* are marked as targeted, non-conditional events.

# Latent TGF- $\beta$ structure and activation

Minlong Shi<sup>1</sup>, Jianghai Zhu<sup>1</sup>, Rui Wang<sup>1</sup>, Xing Chen<sup>1</sup>, Lizhi Mi<sup>1</sup>, Thomas Walz<sup>2</sup> & Timothy A. Springer<sup>1</sup>

**Transforming growth factor (TGF)- $\beta$  is stored in the extracellular matrix as a latent complex with its prodomain. Activation of TGF- $\beta$ 1 requires the binding of  $\alpha_v$  integrin to an RGD sequence in the prodomain and exertion of force on this domain, which is held in the extracellular matrix by latent TGF- $\beta$  binding proteins. Crystals of dimeric porcine proTGF- $\beta$ 1 reveal a ring-shaped complex, a novel fold for the prodomain, and show how the prodomain shields the growth factor from recognition by receptors and alters its conformation. Complex formation between  $\alpha_v\beta_6$  integrin and the prodomain is insufficient for TGF- $\beta$ 1 release. Force-dependent activation requires unfastening of a 'straitjacket' that encircles each growth-factor monomer at a position that can be locked by a disulphide bond. Sequences of all 33 TGF- $\beta$  family members indicate a similar prodomain fold. The structure provides insights into the regulation of a family of growth and differentiation factors of fundamental importance in morphogenesis and homeostasis.**

The TGF- $\beta$  family is key to specifying the body plan during metazoan development<sup>1,2</sup>. Members of this family, including nodal, activins, inhibins, bone morphogenetic proteins (BMPs) and growth differentiation factors (GDFs), specify the anterior/posterior and dorsal/ventral axes, endoderm, mesoderm and ectoderm, left-right asymmetry and details of individual organs. TGF- $\beta$ 1, TGF- $\beta$ 2 and TGF- $\beta$ 3 are important in development, wound healing, immune responses and tumour-cell growth and inhibition<sup>1,3</sup>.

Although TGF- $\beta$  synthesis and expression of its receptors are widespread, activation is localized to sites where TGF- $\beta$  is released from latency. TGF- $\beta$  family members are synthesized with large amino-terminal prodomains, which are required for the proper folding and dimerization of the carboxy-terminal growth-factor domain<sup>4</sup>. Despite intracellular cleavage by furin, after secretion, noncovalent association persists between the dimeric growth-factor domain and prodomain of TGF- $\beta$ , and of an increasingly recognized number of other family members. The prodomain is sufficient to confer latency on some family members and it also targets many members for storage in the extracellular matrix, in complex with latent TGF binding proteins (LTBPs) or fibrillins<sup>5,6</sup>.

The prodomains of TGF- $\beta$ 1 and TGF- $\beta$ 3 contain an RGD motif that is recognized by  $\alpha_v$  integrins. Mice with the integrin-binding RGD motif mutated to RGE recapitulate all major phenotypes of TGF- $\beta$ 1-null mice, including multi-organ inflammation and defects in vasculogenesis, thus demonstrating the essential role of integrins in TGF- $\beta$  activation<sup>7</sup>. Among  $\alpha_v$  integrins, the phenotypes of integrin  $\beta_6$ -null and integrin  $\beta_8$ -null mice demonstrate the particular importance of the  $\alpha_v\beta_6$  and  $\alpha_v\beta_8$  integrins for activation of TGF- $\beta$ 1 and TGF- $\beta$ 3 *in vivo*<sup>8,9</sup>.

Integrin binding alone is not sufficient for TGF- $\beta$  activation. Activation by  $\alpha_v\beta_6$  integrin requires incorporation of TGF- $\beta$ 1 into the extracellular matrix, by association with LTBP, and association of the  $\beta_6$  cytoplasmic domain with the actin cytoskeleton<sup>5,10,11</sup>. Furthermore, contractile force is necessary for TGF- $\beta$  activation by myofibroblasts<sup>8</sup>. Thus, tensile force exerted by integrins across the LTBP-prodomain-TGF- $\beta$  complex is hypothesized to change the conformation of the prodomain and to free TGF- $\beta$  for receptor binding<sup>5,8</sup>. Here, we describe the structure of latent TGF- $\beta$ , mechanisms for latency and integrin-dependent activation, and broad implications for the regulation of bioactivity in the TGF- $\beta$  family.

## Crystal structure

The structure of pro-TGF- $\beta$ 1 at 3.05 Å (Fig. 1a–c and Supplementary Table 1) was solved using multi- and single-wavelength anomalous diffraction. Electron density maps (Supplementary Fig. 1) were improved by multi-crystal, multi-domain averaging over four monomers per asymmetric unit. In a ring-like shape, two prodomain arm domains connect at the elbows to crossed 'forearms' formed by the two growth-factor monomers and by prodomain 'straitjacket' elements that surround each growth-factor monomer (Fig. 1a–c and 4a). The centre of the ring contains solvent. The arms come together at the neck, where they are disulphide-linked in a bowtie, and RGD motifs locate to each shoulder (Fig. 1a). On the opposite side of the ring where the straitjacketed forearms cross, LTBP would be linked to straitjacket residue Cys 4, which is mutated to serine in the crystallization construct (Fig. 1a–c).

The arm domain, residues 46–242, has a novel fold<sup>12</sup> with unusual features. Its two anti-parallel, four-stranded  $\beta$ -sheets bear extensive hydrophobic faces but these overlap only partially in the hydrophobic core (Fig. 1a, e). The hydrophobic faces are extended by long meanders between the two sheets and burial by the  $\alpha 2$ ,  $\alpha 3$  and  $\alpha 4$  helices.

$\beta$ -strands  $\beta 8$  and  $\beta 9$  extend on the two-fold pseudo-symmetry axis to link the two arm domains in a bowtie at the neck (Fig. 1a). The bow is tied with reciprocal inter-prodomain disulphide bonds, Cys 194–Cys 196 and Cys 196–Cys 194, and by hydrophobic residues (Fig. 1a, e).

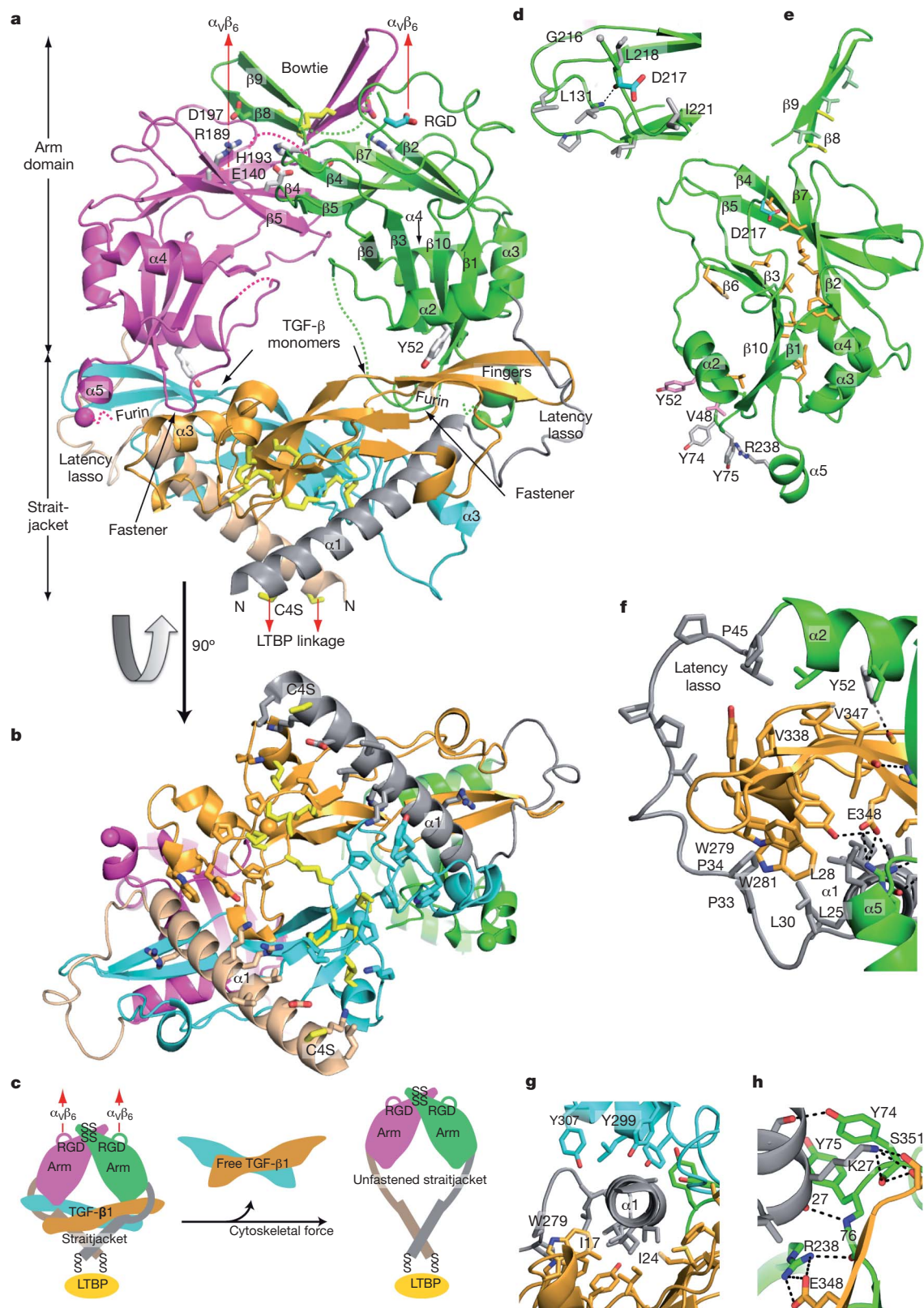
Arg 215 of the RGD motif locates to a disordered loop (residues 209–215) following the bowtie  $\beta 9$  strand. Partially ordered Gly 216 and Asp 217 of the RGD motif (Fig. 1a, d) begin the long, 12-residue meander across the hydrophobic face of the neck-proximal  $\beta$ -sheet that connects to  $\beta 10$  in the forearm-proximal  $\beta$ -sheet.

The straitjacket, residues 1–45, is formed by the  $\alpha 1$  helix and the latency lasso (Fig. 1a–c and 2a). The latency lasso, an extended loop that connects the  $\alpha 1$  and  $\alpha 2$  helices, has little contact with the remainder of the prodomain while encircling the tip of each TGF- $\beta$  monomer (Fig. 1a, b, f). Six proline residues and three aliphatic residues make hydrophobic contacts with an extensive array of growth-factor aromatic and aliphatic residues, and help these to stabilize the conformation of the latency lasso (Fig. 1f).

A highly hydrophobic face of the amphipathic  $\alpha 1$  helix, bearing isoleucine and leucine residues, interacts with Trp 279 and Trp 281 and with aliphatic side chains on one growth-factor monomer (Fig. 1f,

<sup>1</sup>Immune Disease Institute, Children's Hospital Boston and Department of Pathology, Harvard Medical School, Boston, Massachusetts 02115, USA. <sup>2</sup>Howard Hughes Medical Institute and Department of Cell Biology, Harvard Medical School, 240 Longwood Avenue, Boston, Massachusetts 02115, USA.





**Figure 1 | Architecture of proTGF- $\beta$ 1.** Arm, straitjacket and TGF- $\beta$ 1 monomer segments are coloured differently. **a, b**, Overall structure. Spheres mark the last residue visible in density in the prodomain and the first residue of the growth factor. Disordered segments are dashed. Red arrows show the directions of forces during activation by integrins. Key side chains are shown in stick representation, including Asp of the RGD motif in cyan and CED mutations in white. Disulphide bonds and the Cys 4 mutation to Ser are in

yellow. **c**, Schematic of the structure and activation mechanism. SS, disulphide bonds. **d**, Hydrophobic residues near Asp 217 of the RGD motif. **e**, Arm domain. Side chains for the hydrophobic core are shown in gold (also marked in Fig. 2 and Supplementary Fig. 2), conserved  $\alpha$ 2-helix residues that interact with the growth factor are in pink, fastener residues are in silver and bowtie residues are in light green for aliphatics and yellow for Cys. **f–h**, Straitjacket and fastener details.

g). The tryptophan residues are further covered by lasso residues Leu30, Pro33 and Pro34 (Fig. 1f). Notably, the  $\alpha 1$  helix is buried deeply in an interface between the two growth-factor monomers (Fig. 1a, b, g). The interface on the second monomer includes three tyrosine residues (Fig. 1g).

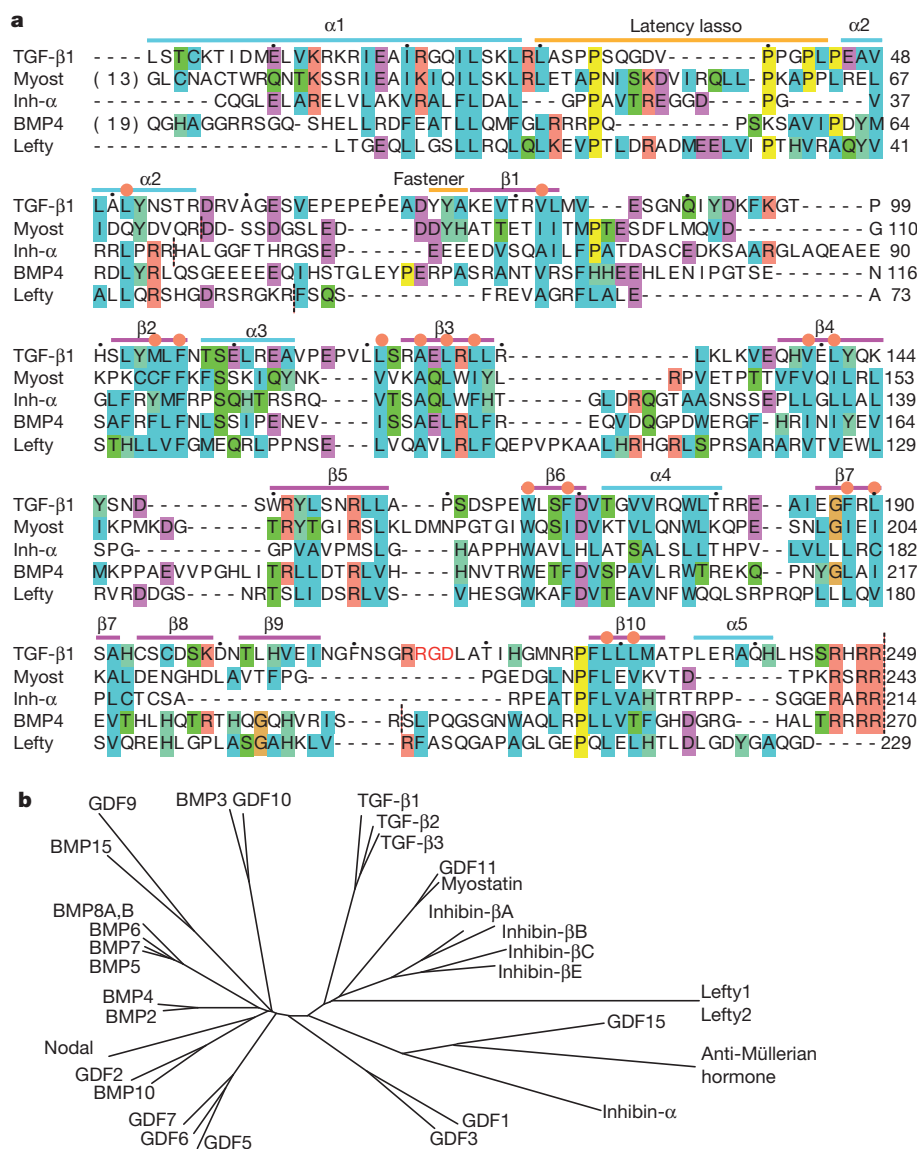
Together with the straitjacket, the arm domain completes the encirclement of each growth-factor monomer. The  $\alpha 2$  helix buries Val 338 and Val 347 of the TGF- $\beta$  finger (Fig. 1f). The  $\alpha 5$  helix projects from the base of the arm domain (Fig. 1a and e) and in it, prodomain residue Arg 238 forms a salt bridge to Glu 348 in TGF- $\beta 1$  (Fig. 1f, h). These two residues are invariant in proTGF- $\beta 1$ , proTGF- $\beta 2$  and proTGF- $\beta 3$  (Supplementary Fig. 2).

The straitjacket is fastened to arm residues 74–76 (Fig. 1a, h). A backbone hydrogen bond between the nitrogen of Ala 76 and the oxygen of Lys 27 caps the C-terminal end of the  $\alpha 1$  helix (Fig. 1h). Moreover, the carbonyl oxygen of Ala 76 forms a hydrogen bond to Arg 238 in the  $\alpha 5$  helix (Fig. 1h). Lys 27 is a key fastener residue. Its side chain forms a  $\pi$ -cation bond to the side chain of Tyr 74, a hydrogen bond to the backbone of Tyr 74, and hydrogen bonds to the backbone and sidechain of Ser 351 (Fig. 1h). Van der Waals contacts between the bulky side chains of the fastener residues Lys 27, Tyr 74

and Tyr 75 also secure the straitjacket. Notably, Lys 27, Tyr 74 and Tyr 75 are invariant among TGF- $\beta 1$ , TGF- $\beta 2$  and TGF- $\beta 3$  (Supplementary Fig. 2). Fastening is reinforced by backbone hydrogen bonds between arm  $\beta 1$ -strand residues 77 and 78 and growth-factor  $\beta$ -fingers, which join the prodomain and the growth factor in a super- $\beta$ -sheet (Fig. 1a).

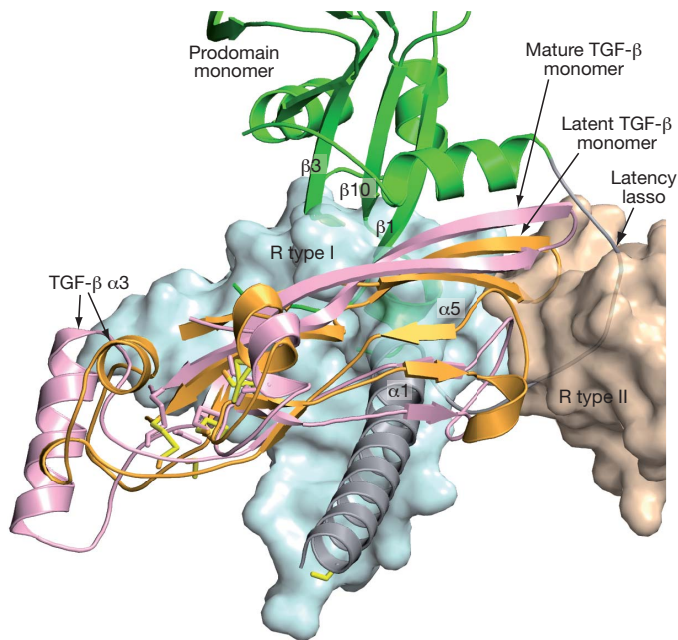
The TGF- $\beta$  dimer forms the forearms, although TGF- $\beta$  monomers have also been described as hand-like<sup>13–15</sup> (Fig. 3). Each monomer has no hydrophobic core, aside from the cystine knot motif in which one disulphide passes between two polypeptide segments bridged by two other disulphides.

Prodomain-bound TGF- $\beta 1$  differs markedly from previous TGF- $\beta$  structures in both the orientation between monomers and the position of elements within monomers (Fig. 3 and Supplementary Fig. 3). The C $\alpha$  root-mean-squared deviations over all 112 residues are 7 Å, and 2 Å over the most similar 85 residues. The largest differences are imposed by the prodomain  $\alpha 1$  helix. It occupies a similar position to the growth-factor  $\alpha 3$  helix in mature TGF- $\beta 1$  (Supplementary Fig. 3). Intercalation of the prodomain  $\alpha 1$  helix between the growth-factor monomers reduces the total area buried between monomers from 850 Å<sup>2</sup> to 335 Å<sup>2</sup>. The large conformational changes in TGF- $\beta 1$  are



**Figure 2 | The TGF- $\beta$  family.** **a**, Sequence alignment of five representative prodomains. Orange circles mark the core hydrophobic residues shown in Fig. 1e. Inh- $\alpha$ , inhibin- $\alpha$ ; Myost, myostatin. Black dots over TGF- $\beta 1$  sequence

mark decadal residues. Vertical dashed lines mark cleavage sites. **b**, Phylogenetic tree of the TGF- $\beta$  family, based on the alignment in Supplementary Fig. 2.



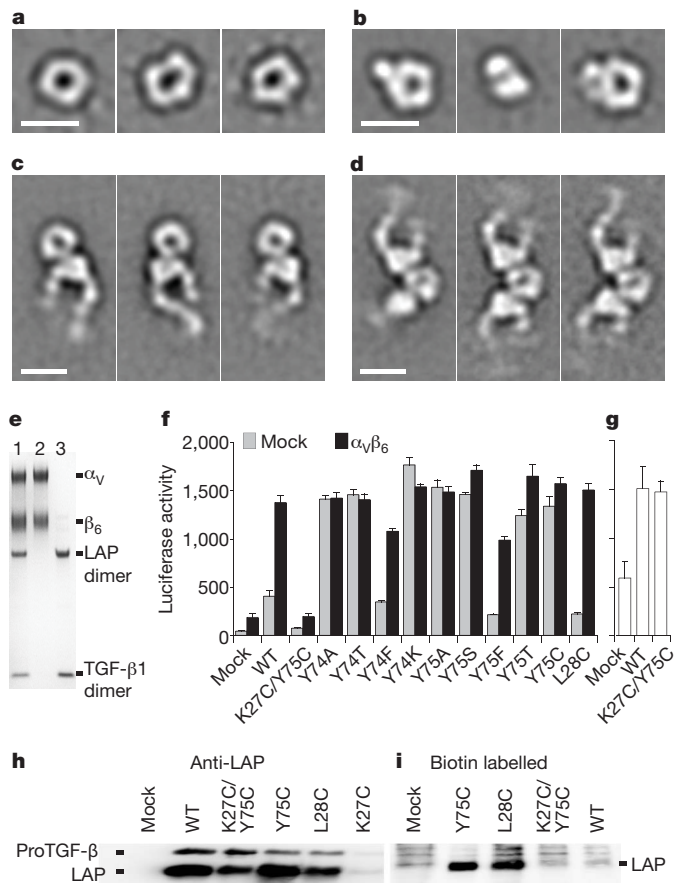
**Figure 3 | Shielding from receptor binding.** ProTGF-β1 and TGF-β1 in complex with its receptors (R type I and II) (ref. 15) were superimposed on the TGF-β dimers. For clarity, only one monomer of each is shown. The receptors are shown as transparent molecular surfaces. Elements of the prodomain that clash with the receptors are labelled.

driven by an intimate interaction between the growth-factor and prodomain dimers, which buries a total area of 2,440 Å<sup>2</sup>.

### Implications for biosynthesis

Folding and secretion of active TGF-β1 and activin A requires the co-expression of their proddomains<sup>4</sup>, whereas the TGF-β1 prodomain can be biosynthesized in the absence of the growth-factor domain<sup>16</sup>. These findings suggest that the C-terminal growth-factor domain folds either concomitantly with, or subsequently to, the N-terminal prodomain. The kinetics of biosynthesis of TGF-β1, activin and anti-Müllerian hormone are slow and for anti-Müllerian hormone, folding of the growth-factor domain is rate-limiting<sup>17</sup>. Regions of the prodomain that may be particularly important in templating the folding of the growth-factor domain include the β1 strand that forms a supersheet with the TGF-β fingers and the α1 and α2 helices, which pack against extensive hydrophobic interfaces on opposite sides of the growth-factor fingers (Fig. 1a, f, g). Residues Ile 17, Ile 24, Leu 25 and Leu 28 in the α1-helix interface, and Leu 30 in the lasso interface (Fig. 1f, g), have been specifically identified as important for TGF-β1 association<sup>18</sup>. The embrace of the fingers of each growth-factor monomer may complement the correct formation of the cystine knot and inter-monomer disulphide bonds in TGF-β. The structure of proTGF-β1 makes these disulphides accessible to disulphide isomerases during biosynthesis (Fig. 1b).

A definitive assignment of which growth-factor and prodomain monomers derive from the same polypeptide chain is not possible because of intracellular cleavage by furin and lack of density for residues 243–249. However, cleavage is incomplete and the small amount of uncleaved proTGF-β that is present in protein preparations co-crystallizes with cleaved proTGF-β (Supplementary Fig. 1d), indicating that there is no major conformational change after cleavage. A long prodomain–growth-factor connection through the centre of the ring, spanning ~50 Å, would require substantial conformational change and would limit access to furin. Therefore, we have assigned the shorter ~30 Å connection, between the C terminus of the prodomain and the N terminus of the growth factor, located on the same side of the ring (for example, the magenta and gold spheres in Fig. 1a, b).



**Figure 4 | ProTGF-β1 complexes with LTBP and α<sub>v</sub>β<sub>6</sub> integrin, and activation of TGF-β.** **a–d**, Representative negative-stain electron microscopy class averages of proTGF-β (**a**), the complex of proTGF-β with a fragment of LTBP1 containing TGF-binding domain 3–EGF–EGF–TGF-binding domain 4 (**b**) and complexes of proTGF-β1 with α<sub>v</sub>β<sub>6</sub> integrin, prepared with an excess of proTGF-β1 (**c**) or an excess of α<sub>v</sub>β<sub>6</sub> integrin (**d**). Scale bars, 100 Å. **e**, Non-reducing SDS–PAGE of the complex peak from S200 gel filtration used for electron microscopy in **d** (lane 1), α<sub>v</sub>β<sub>6</sub> integrin (lane 2) and proTGF-β (lane 3). LAP, latency-associated protein (prodomain). **f**, Activation of TGF-β1. 293T cells stably transfected with α<sub>v</sub>β<sub>6</sub> integrin or a mock control were additionally transfected with the indicated wild-type (WT) or mutant proTGF-β1 constructs, or with empty vector (mock), and co-cultured with TGF-β indicator cells<sup>5</sup>. **g**, Material made by the indicated mutants in 293T cells was heated at 80 °C and assayed with indicator cells. Error bars in **f** and **g** show the s.e.m. of 3–9 samples from 1–3 representative experiments. **h**, **i**, Western blots of proTGF-β1 secreted by the indicated transfectants, using an antibody to the prodomain (**h**) or streptavidin to detect biotinylated cysteines (**i**).

This assignment indicates a swap in the growth-factor monomer that each prodomain monomer embraces, with the intimate interactions described above occurring between proddomains and growth-factor domains that are present on different polypeptide chains in the precursor in the endoplasmic reticulum (for example, the gold and green domains in Fig. 1). Thus, the surface area buried between the growth-factor domains and proddomains of different precursor monomers (900 Å<sup>2</sup>) is substantially larger than that within the same precursor monomer (370 Å<sup>2</sup>) and adds substantially to the inter-monomer interfaces of the growth factor (330 Å<sup>2</sup>) and prodomain (600 Å<sup>2</sup>). Swapping may be important in regulating the proportion of growth-factor heterodimers, which have unique functions in settings such as dorsoventral patterning<sup>19</sup>.

### Complex with LTBP and activation

In the large latent complex, a single LTBP molecule is disulphide-bonded to two proTGF-β monomers. We confirmed this unusual 1:2



stoichiometry by multi-angle light-scattering mass measurements of the complex with an LTBP fragment. The LTBP-crosslinking Cys 4 residues in each monomer (serines in our construct) are 40 Å apart (Fig. 1b). Their linkage to LTBP will reinforce the straitjacket by fastening together the forearms (Fig. 1c). The large 40 Å separation may represent a mechanism for preventing disulphide linkage between the two Cys 4 residues. In contrast, the two cognate cysteines in TGF- $\beta$ -binding domain 3 of LTBP become linked to one another in the absence of complex formation. These cysteines are surface-exposed and surrounded by acidic residues that interact with proTGF- $\beta$  (refs 20, 21). In agreement with this, the straitjacket  $\alpha$ 1-helices bear basic residues (Fig. 1b). Moreover, between the two prodomain  $\alpha$ 1-helices, a concave growth-factor surface that bears numerous hydrophobic residues, including proline and disulphide-linked cysteine, is available for interaction with LTBP (Fig. 1b).

Negative-stain electron microscopy class averages of proTGF- $\beta$  are in excellent agreement with our crystal structure (Fig. 4a and Supplementary Fig. 4). A complex with an LTBP fragment that contains TGF- $\beta$ -binding domains 3 and 4 and two intervening EGF domains shows no major conformational change in the proTGF- $\beta$  moiety (Fig. 4b). An additional density corresponding to the LTBP fragment is present on the periphery of the ring, as expected from our crystal structure, and causes the ring to lie at an angle to the substrate in some class averages (Fig. 4b, middle panel and Supplementary Fig. 4c).

The RGD motifs in the shoulders of proTGF- $\beta$  are highly accessible for integrin binding (Fig. 1a). In contrast to many RGD motifs that are present in extended loops, the position of Asp 217 is stabilized by burial of Leu 218 and by a 218–131 backbone hydrogen bond (Fig. 1d). Exposed hydrophobic side chains that are nearby on the body of the arm domain (Fig. 1d) may increase affinity for integrins.

The integrin  $\alpha_v\beta_6$  ectodomain, with a C-terminal clasp, formed complexes with proTGF- $\beta$ 1 in  $\text{Ca}^{2+}$  and  $\text{Mg}^{2+}$  that could be isolated by gel filtration, demonstrating unusually tight binding for an integrin (Fig. 4c–e and Supplementary Fig. 4a). Different input ratios of proTGF- $\beta$ 1 and  $\alpha_v\beta_6$  integrin yielded 1:1 (Fig. 4c) and 1:2 (Fig. 4d, e) complexes. Binding to ligand stabilized extension of the integrin legs and the open conformation of the  $\alpha_v\beta_6$  headpiece<sup>22</sup> (Fig. 4c, d and Supplementary Fig. 4d, e). Integrins were bound to proTGF- $\beta$  at the interface between a large density, corresponding to the  $\alpha_v$   $\beta$ -propeller domain, and a small density, corresponding to the  $\beta_6$   $\beta$ 1 domain, with their legs extending away from proTGF- $\beta$ . The spacing between the binding sites on the ring seen by electron microscopy (40–50 Å) was appropriate for that between the two RGDs in the crystal structure (45 Å). No major conformational change in proTGF- $\beta$ 1 was apparent, even with two integrins bound (Fig. 4d), and SDS–polyacrylamide gel electrophoresis (SDS–PAGE) confirmed the presence of TGF- $\beta$ 1 in the complex (Fig. 4e). These biochemical and structural studies demonstrate that integrin binding to proTGF- $\beta$ 1 is not sufficient for release of TGF- $\beta$ 1, consistent with previous cell-biological assays<sup>5,8,10</sup>. The requirements of (1) attachment of proTGF- $\beta$  through LTBP to the extracellular matrix; (2) integrin attachment to the cytoskeleton and (3) cellular contraction indicate that the generation of tensile force across proTGF- $\beta$ 1 is required for activation of TGF- $\beta$ <sup>5,8,10,11</sup>.

The structure enables the overall mechanism of TGF- $\beta$ 1 activation by applied force to be readily predicted (Fig. 1c). Tensile force applied to the RGD motifs in the shoulders by  $\alpha_v$  integrins attached to the actin cytoskeleton will be resisted at the opposite end of the ring, where the Cys 4 residues in the straitjacket are disulphide-linked to LTBP, which is tightly associated with the extracellular matrix. Pulling force will be applied in the directions shown by red arrows in Fig. 1a.

The direction of the pulling force and fold topology strongly influence the unfolding pathway and resistance to force<sup>23</sup>.  $\beta$ -Sheet proteins are the most force-resistant and thus the arm domain will be the most force-resistant portion of the prodomain. Pulling against the RGD motif will be transmitted through the long meander to the  $\beta$ 10 strand. Force transmitted from the Cys 4 residues through the straitjacket will

be resisted by the  $\beta$ 1 strand. The  $\beta$ 1 and  $\beta$ 10 strands are each parallel to applied force and adjacent in a  $\beta$ -sheet (Fig. 1) and are thus in the most force-resistant structural geometry known, the hydrogen-bond clamp<sup>23</sup>. By contrast, the topologies and geometries of the  $\alpha$ -helices and the long latency lasso of the straitjacket are ill-suited to resist force. Force on Cys 4 will apply leverage to the C-terminal end of the  $\alpha$ 1 helix and weaken interactions with fastener residues. After unfastening, the long latency lasso, which has no stabilizing hydrogen-bond interactions, will be easily elongated and straightened by the applied tensile force. Thus, freed by opening of its straitjacket, TGF- $\beta$  will be released from the prodomain and activated for receptor binding (Fig. 1c).

The prodomain not only holds TGF- $\beta$  in a markedly different conformation from when it is free or bound to receptors, it also blocks receptor access completely. TGF- $\beta$  family members are recognized by two type I receptors and two type II receptors that surround the growth-factor dimer (Supplementary Fig. 3e)<sup>15</sup>. Binding of the type II receptor to the finger-tips of the growth factor is blocked by the latency lasso, and binding of the type I receptor to the body of the growth-factor domain is blocked by the prodomain  $\alpha$ 1 helix,  $\alpha$ 5 helix, the fastener and the ends of  $\beta$ -strands 1, 3 and 10 (Fig. 3). Although straitjacket removal might be sufficient to allow binding of type II receptors, type I receptor interactions overlap with so many interactions between TGF- $\beta$  and the arm domain (Fig. 3) that complete release from the prodomain would be required for receptor binding. The structure thus shows that integrins could not expose TGF- $\beta$  sufficiently for receptor activation if it remained bound to the prodomain, and that other explanations should be sought for the greater activity of integrin-activated TGF- $\beta$  on neighbouring cells than on distant cells<sup>10</sup>.

To test the importance of unfastening in TGF- $\beta$  activation, key residues were mutated. Non-conservative substitutions of the fastener residues Tyr 74 and Tyr 75 resulted in spontaneous, non-integrin-dependent TGF- $\beta$  activation (Fig. 4f). Among the different amino acids to which Tyr 74 and Tyr 75 were mutated, only phenylalanine was not activating. As a control, mutation of nearby Leu 28, in a hydrophobic interface with TGF- $\beta$ , was not activating (Fig. 4f). These results are consistent with the importance to fastening of  $\pi$ -bonding and of van der Waals interactions of the aromatic tyrosine side chains.

In the fastener, the C $\alpha$  carbons of Lys 27 and Tyr 75 are only 4.1 Å apart, permissive for disulphide bond formation in a K27C/Y75C mutant, as confirmed by free-cysteine labelling of the Y75C mutant but not the K27C/Y75C mutant (Fig. 4h, i). The K27C mutation greatly reduced expression (Fig. 4h). Similarly, a K27A mutation greatly reduces expression, and also releases free TGF- $\beta$ 1 (ref. 18). The Y75C mutant was constitutively active (Fig. 4f). The K27C/Y75C double mutation rescued expression compared to K27C, prevented the spontaneous release of TGF- $\beta$  that was seen with Y75C and, compared to wild type, made proTGF- $\beta$  completely resistant to integrin- $\alpha_v\beta_6$ -dependent activation (Fig. 4f, h). Denaturants such as heat can unfold proteins by pathways distinct from applied force<sup>23</sup>. Heat released comparable amounts of active TGF- $\beta$  from wild type and the K27C/Y75C mutant (Fig. 4g). Thus, a disulphide bond can fasten the straitjacket permanently and prevent integrin-dependent activation. These results support the hypothesis that tensile force applied to the prodomain by integrins can release TGF- $\beta$ , and emphasise the importance of straitjacket unfastening in integrin-dependent activation.

## Mutations in disease

Camurati–Engelmann disease (CED), which is characterized by thickening of the shafts of the long bones with pain in muscle and bone, is caused by mutations in the prodomain of TGF- $\beta$ 1 that increase its release<sup>18,24</sup>. Among CED mutations, Y52H disrupts an  $\alpha$ 2-helix residue that cradles the TGF- $\beta$  fingers (Fig. 1a, f). The charge-reversal E140K and H193D mutations disrupt a pH-regulated

salt bridge between Glu 140 and His 193 in the dimerization interface of the prodomain (Fig. 1a). 'Hotspot' residue Arg 189 is substantially buried: it forms a cation- $\pi$  bond with Tyr 142 and salt bridges across the dimer interface with bowtie residue Asp 197 (Fig. 1a). Moreover, CED mutations in Cys 194 and Cys 196 demonstrate the importance of the bowtie disulphide bonds.

### Implications for the large TGF- $\beta$ family

The TGF- $\beta$  family consists of 33 members (Fig. 2b)<sup>2</sup>. Although growth-factor domains are highly conserved, prodomains vary in length from 169 to 433 residues, and are variously described as unrelated in sequence or low in homology. However, alignment shows that all prodomains have a similar fold (Fig. 2a and Supplementary Fig. 2). Deeply buried hydrophobic residues in core secondary-structure elements of the arm domain, that is, the  $\alpha$ 2 helix and  $\beta$ -strands 1–3, 6, 7 and 10, are conserved in all members (gold side chains in Fig. 1e and orange circles in Fig. 2 and Supplementary Fig. 2).

Most family members also contain clear sequence signatures for the amphipathic C-terminal portion of the  $\alpha$ 1 helix that inserts intimately between the two growth-factor monomers (Fig. 2 and Supplementary Fig. 2). A similar insertion in inhibin- $\alpha$  and inhibin- $\beta$ A has been demonstrated by mapping disruptive mutations to the equivalents of Ile 24 and Leu 28 in TGF- $\beta$  (Fig. 1f, g)<sup>25</sup>. Many family members also contain proline-rich latency lasso loops with lengths that are compatible with encirclement of the growth-factor  $\beta$ -finger (Fig. 2 and Supplementary Fig. 2). Thus, a prodomain structure similar to that of proTGF- $\beta$ , including a portion of the straitjacket, is widespread in the TGF- $\beta$  family. However, the low sequence identity and many insertions and deletions indicate substantial specializations.

Differences in prodomain dimerization among family members are indicated by variations in cysteine positions. The bowtie ( $\beta$ -strands 8 and 9) and its disulphides are specializations. Inhibin- $\alpha$  and - $\beta$  subunits have cysteines in similar positions, whereas other family members either have cysteine residues in the  $\beta$ 7 strand or lack cysteines altogether in this region (Fig. 2 and Supplementary Fig. 2).

The interface between the two arm domains in the  $\beta$ 4 and  $\beta$ 5 strands is modest in size and lacks hydrophobic and conserved residues. GDF1 and GDF15 specifically lack the  $\beta$ 4 and  $\beta$ 5 strands, which are adjacent in sequence and structure, on the edge of a  $\beta$ -sheet (Fig. 1a and Supplementary Fig. 2). Therefore, arm-domain dimerization seems to be variable or absent in some family members.

The close relatives of TGF- $\beta$ , myostatin and GDF11, which are also latent, show conservation of the fastener residues Lys 27 and Tyr 75 (Fig. 2a and Supplementary Fig. 2). Myostatin regulates muscle mass and is stored in the extracellular matrix, bound to LTBP3. Release of myostatin and GDF11 from latency requires cleavage of the prodomain between Arg 75 and Asp 76 by BMP1/tolloid metalloproteinases (reviewed in ref. 26). This cleavage is between the  $\alpha$ 2 helix and the fastener (Fig. 2a). Thus at least two different methods of unfastening the straitjacket, force and proteolysis, can release family members from latency.

An increasingly large number of TGF- $\beta$  family members are recognized to remain associated with their prodomains after secretion, including BMP4, BMP7, BMP10, GDF2, GDF5 and GDF8 (ref. 27). Furthermore, many of these prodomains bind with high affinity to fibrillin-1 and fibrillin-2. Targeting by the prodomain to the extracellular matrix may be of wide importance in regulating bioactivity in the TGF- $\beta$  family<sup>6</sup>. Moreover, binding to LTBP3 or fibrillins seems to strengthen the prodomain-growth-factor complex<sup>6</sup>. Thus, although only a limited number of TGF- $\beta$  family members are latent as prodomain-growth-factor complexes, the concept of latency may extend to other members when their physiologically relevant complexes with LTBP3 and fibrillins are considered.

The signalling range of BMP4 *in vivo* is increased by extracellular cleavage of the prodomain by furin-like proteases at a second site upstream of the prodomain-growth-factor cleavage site<sup>28</sup>. Notably,

the second site is in the disordered loop bearing the arginine of RGD in TGF- $\beta$ 1 (Fig. 2a). Loss of the central  $\beta$ 10 strand between the two cleavage sites results in loss of binding of the BMP4 prodomain to its growth factor<sup>27</sup>.

The prodomain of Nodal, which binds to Cripto, targets Nodal for cleavage by proteases secreted by neighbouring cells<sup>29</sup>. Anti-Müllerian hormone is secreted largely uncleaved and association with the prodomain greatly potentiates its activity *in vivo*<sup>30</sup>. Lefty protein, which is involved in establishing bilateral asymmetry, is not cleaved between the arm and growth-factor domains, and is cleaved instead between the  $\alpha$ 2 helix and the fastener<sup>31</sup> (Fig. 2). Notably, release of the straitjacket should be sufficient to enable access of type II receptors to growth-factor domains.

### Concluding perspective

We have described the structure of latent TGF- $\beta$ 1 and a force-dependent mechanism for its activation by  $\alpha_v$  integrins. It is notable that so many members of the TGF- $\beta$  family associate with fibrillins or with LTBP3, which co-assemble in the elastic fibres of connective tissues<sup>6</sup>. Forces acting on elastic fibres would extend fibrillins and LTBP3, and we speculate that this could weaken their association with TGF- $\beta$  family members, enabling release and activation. It is thus possible that force-dependent regulation of TGF- $\beta$  family activation could extend beyond integrin-dependent mechanisms and could be important in a wide variety of contexts, including regulation of bone and tissue growth. Although prodomains in the TGF- $\beta$  family are diverse, their sequences are highly conserved between species<sup>2</sup>. Further studies are required to address the diversity of mechanisms by which prodomains regulate latency and activation in the TGF- $\beta$  family.

### METHODS SUMMARY

Porcine proTGF- $\beta$ 1 with an N-terminal tag, C4S and N147Q mutations was expressed in CHO-Lec 3.2.8.1 cells. The protein was purified, treated with 3C protease to remove tags and then crystallized. Diffraction-quality crystals were obtained in 6.5–7.5% (w/v) polyethylene glycol 3500, 17–18% isopropanol, 4–5% glycerol and 0.1 M sodium citrate, pH 5.6. Structures were solved by Se multi-wavelength anomalous dispersion and Hg single-wavelength anomalous dispersion. Maps were improved by multi-crystal averaging. The structure was built manually and refined to  $R_{\text{work}}$  and  $R_{\text{free}}$  factors of 27.4% and 31.1%, respectively. The  $\alpha_v\beta_6$  ectodomain was expressed using C-terminal  $\alpha$ -helical coiled-coils and tags, purified and subjected to negative-stain electron microscopy. TGF- $\beta$  assays used 293T cells stably transfected with  $\alpha_v\beta_6$  integrin and transiently transfected with mutant human proTGF- $\beta$ 1, then co-cultured with indicator cells expressing a luciferase gene under the control of a TGF- $\beta$ 1-inducible promoter.

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Received 20 December 2010; accepted 27 April 2011.**

1. Wu, M. Y. & Hill, C. S. TGF- $\beta$  superfamily signaling in embryonic development and homeostasis. *Dev. Cell* **16**, 329–343 (2009).
2. Derynck, R. & Miyazono, K. In *The TGF- $\beta$  Family* (eds Derynck, R. & Miyazono, K.) Ch. 2, 29–43 (Cold Spring Harbor Laboratory Press, 2008).
3. Blobbe, G. C., Schiemann, W. P. & Lodish, H. F. Role of transforming growth factor  $\beta$  in human disease. *N. Engl. J. Med.* **342**, 1350–1358 (2000).
4. Gray, A. M. & Mason, A. J. Requirement for activin A and transforming growth factor- $\beta$  1 pro-regions in homodimer assembly. *Science* **247**, 1328–1330 (1990).
5. Annes, J. P., Chen, Y., Munger, J. S. & Rifkin, D. B. Integrin  $\alpha_v\beta_6$ -mediated activation of latent TGF- $\beta$  requires the latent TGF- $\beta$  binding protein-1. *J. Cell Biol.* **165**, 723–734 (2004).
6. Ramirez, F. & Sakai, L. Y. Biogenesis and function of fibrillin assemblies. *Cell Tissue Res.* **339**, 71–82 (2010).
7. Yang, Z. *et al.* Absence of integrin-mediated TGF $\beta$ 1 activation *in vivo* recapitulates the phenotype of TGF $\beta$ 1-null mice. *J. Cell Biol.* **176**, 787–793 (2007).
8. Wipff, P. J. & Hinz, B. Integrins and the activation of latent transforming growth factor  $\beta$  1 — an intimate relationship. *Eur. J. Cell Biol.* **87**, 601–615 (2008).
9. Aluwihare, P. *et al.* Mice that lack activity of  $\alpha_v\beta_6$ - and  $\alpha_v\beta_5$ -integrins reproduce the abnormalities of *Tgfb1*- and *Tgfb3*-null mice. *J. Cell Sci.* **122**, 227–232 (2009).
10. Munger, J. S. *et al.* The integrin  $\alpha_v\beta_6$  binds and activates latent TGF  $\beta$ 1: A mechanism for regulating pulmonary inflammation and fibrosis. *Cell* **96**, 319–328 (1999).

11. Yoshinaga, K. *et al.* Perturbation of transforming growth factor (TGF)- $\beta$ 1 association with latent TGF- $\beta$  binding protein yields inflammation and tumors. *Proc. Natl Acad. Sci. USA* **105**, 18758–18763 (2008).
12. Holm, L., Kaariainen, S., Rosenstrom, P. & Schenkel, A. Searching protein structure databases with DaliLite v.3. *Bioinformatics* **24**, 2780–2781 (2008).
13. Daopin, S., Piez, K. A., Ogawa, Y. & Davies, D. R. Crystal structure of transforming growth factor- $\beta$ 2: an unusual fold for the superfamily. *Science* **257**, 369–373 (1992).
14. Schlunegger, M. P. & Grutter, M. G. An unusual feature revealed by the crystal structure at 2.2 Å resolution of human transforming growth factor- $\beta$ 2. *Nature* **358**, 430–434 (1992).
15. Radaev, S. *et al.* Ternary complex of transforming growth factor- $\beta$ 1 reveals isoform-specific ligand recognition and receptor recruitment in the superfamily. *J. Biol. Chem.* **285**, 14806–14814 (2010).
16. Gentry, L. E. & Nash, B. W. The pro domain of pre-pro-transforming growth factor  $\beta$ 1 when independently expressed is a functional binding protein for the mature growth factor. *Biochemistry* **29**, 6851–6857 (1990).
17. Belville, C. *et al.* Mutations of the anti-Müllerian hormone gene in patients with persistent Müllerian duct syndrome: biosynthesis, secretion, and processing of the abnormal proteins and analysis using a three-dimensional model. *Mol. Endocrinol.* **18**, 708–721 (2004).
18. Walton, K. L. *et al.* Two distinct regions of latency-associated peptide coordinate stability of the latent transforming growth factor- $\beta$ 1 complex. *J. Biol. Chem.* **285**, 17029–17037 (2010).
19. Little, S. C. & Mullins, M. C. Bone morphogenetic protein heterodimers assemble heteromeric type I receptor complexes to pattern the dorsoventral axis. *Nature Cell Biol.* **11**, 637–643 (2009).
20. Lack, J. *et al.* Solution structure of the third TB domain from LTBP1 provides insight into assembly of the large latent complex that sequesters latent TGF- $\beta$ . *J. Mol. Biol.* **334**, 281–291 (2003).
21. Chen, Y. *et al.* Amino acid requirements for formation of the TGF- $\beta$ -latent TGF- $\beta$  binding protein complexes. *J. Mol. Biol.* **345**, 175–186 (2005).
22. Luo, B.-H., Carman, C. V. & Springer, T. A. Structural basis of integrin regulation and signaling. *Annu. Rev. Immunol.* **25**, 619–647 (2007).
23. Forman, J. R. & Clarke, J. Mechanical unfolding of proteins: insights into biology, structure and folding. *Curr. Opin. Struct. Biol.* **17**, 58–66 (2007).
24. Janssens, K. *et al.* Camurati-Engelmann disease: review of the clinical, radiological, and molecular data of 24 families and implications for diagnosis and treatment. *J. Med. Genet.* **43**, 1–11 (2005).
25. Walton, K. L. *et al.* A common biosynthetic pathway governs the dimerization and secretion of inhibin and related transforming growth factor  $\beta$  (TGF $\beta$ ) ligands. *J. Biol. Chem.* **284**, 9311–9320 (2009).
26. Anderson, S. B., Goldberg, A. L. & Whitman, M. Identification of a novel pool of extracellular pro-myostatin in skeletal muscle. *J. Biol. Chem.* **283**, 7027–7035 (2008).
27. Sengle, G. *et al.* Targeting of bone morphogenetic protein growth factor complexes to fibrillin. *J. Biol. Chem.* **283**, 13874–13888 (2008).
28. Cui, Y. *et al.* The activity and signaling range of mature BMP-4 is regulated by sequential cleavage at two sites within the prodomain of the precursor. *Genes Dev.* **15**, 2797–2802 (2001).
29. Blanchet, M. H. *et al.* Cripto recruits Furin and PACE4 and controls Nodal trafficking during proteolytic maturation. *EMBO J.* **27**, 2580–2591 (2008).
30. Wilson, C. A. *et al.* Müllerian inhibiting substance requires its N-terminal domain for maintenance of biological activity, a novel finding within the transforming growth factor- $\beta$  superfamily. *Mol. Endocrinol.* **7**, 247–257 (1993).
31. Ulloa, L. *et al.* Lefty proteins exhibit unique processing and activate the MAPK pathway. *J. Biol. Chem.* **276**, 21387–21396 (2001).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank P. Sun for porcine TGF- $\beta$ 1 cDNA, K. Koli for human TGF- $\beta$ 1 and LTBP1 cDNAs, D. Rifkin for human LTBP1 cDNA and transformed mink lung epithelial cells, and the staff of the Advanced Photon Source General Medical Sciences and National Cancer Institute (APS GM/CA-CAT) beamline 23-ID.

**Author Contributions** M.S. cloned, expressed and purified proTGF- $\beta$ 1, crystallized the protein, collected and processed X-ray data, refined and analysed the structure, designed and performed biochemical assays and wrote the paper. J.Z. collected and processed X-ray data, refined and analysed the structure and performed electron microscopy studies. R.W. designed and performed TGF- $\beta$ 1 assays. X.C. performed electron microscopy studies. L.M. processed X-ray data. T.W. provided electron microscopy supervision. T.A.S. designed and supervised the project, refined and analysed the structure and wrote the paper.

**Author Information** X-ray structures have been deposited in the Protein Data Bank under the accession number 3RJR. Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at [www.nature.com/nature](http://www.nature.com/nature). Correspondence and requests for materials should be addressed to T.A.S. ([springer@idi.harvard.edu](mailto:springer@idi.harvard.edu)).



## METHODS

**ProTGF- $\beta$ 1.** The porcine proTGF- $\beta$ 1 construct with the rat serum albumin leader sequence (MKWVTFLLLLFISGSFAFS), followed by eight histidine residues, a streptavidin-binding peptide (TTGWRGGHVVVELAGELEQLRARLEHH PQGQREP)<sup>32</sup> and a HRV-3C protease site (LEVLFGQP) was amplified from pcDNA-GS-TGF- $\beta$ 1 (ref. 33). Porcine proTGF- $\beta$ 1 with the C4S mutation was amplified from the latter construct. The C4S mutation increases proTGF- $\beta$ 1 expression<sup>33</sup> and avoids inappropriate disulphide bond formation<sup>34</sup>. No crystals were obtained with this construct. One or two N-linked sites were deleted in the N147Q and N107Q/N147Q constructs, which were expressed similarly to wild type<sup>35</sup>. The best crystals were obtained with N147Q; the N107Q/N147Q mutant yielded needles that could not be optimized. CHO-Lec 3.2.8.1 cells were transfected by electroporation and cultured with 10  $\mu$ g ml<sup>-1</sup> puromycin. Clones were screened for expression using a sandwich enzyme-linked immunosorbent assay (ELISA) with a capture antibody to prodomain-1 (R&D Systems) and a biotinylated detection antibody to the His tag (Qiagen). The clone with highest expression of proTGF- $\beta$ 1 ( $\sim$ 2 mg l<sup>-1</sup>) was expanded and cultured in roller bottles with J/J medium and 5% fetal bovine serum (FBS). Supernatants were collected every 5 d, clarified by centrifugation, concentrated tenfold with tangential flow filtration (Vivaflow 200, Sartorius Stedim), diluted fivefold with 10 mM Tris-HCl, 0.14 M NaCl (TBS, pH 8.0), then concentrated fivefold. Material was adjusted to 0.2 M NaCl and purified using Ni-NTA agarose (Qiagen) (25 ml per 5 l of culture supernatant), then washed with three column volumes of 0.6 M NaCl, 0.01 M Tris (pH 8.0) and eluted with 0.25 M imidazole in TBS. Material was adjusted to pH 7.4, applied to Strep-tactin agarose (IBA) (3 ml per 5 l of culture supernatant) and washed with TBS (pH 7.4). Then 4 ml of recombinant His-tagged HRV-3C protease (Novagen, 100 U mg<sup>-1</sup>, 1 mg ml<sup>-1</sup>), diluted 20-fold in TBS (pH 7.4) with 10% glycerol, was applied to the column, which was held at 4 °C for 16 h. The flow-through with two column volumes of TBS (pH 7.4), containing untagged proTGF- $\beta$ 1, was concentrated to 1 ml and applied to MonoQ and Superdex S200 columns connected in series and equilibrated with TBS (pH 7.5). Purified proTGF- $\beta$ 1 was concentrated to about 15 mg ml<sup>-1</sup> in 10 mM Tris (pH 7.5), 75 mM NaCl for crystal screening in 96-well Greiner microplates (100 nl hanging-drop vapour diffusion format) using a mosquito crystallization robot (Molecular Dimensions) at 20 °C. Hits were optimized in 24-well plates using hanging-drop vapour diffusion. However, better-diffracting crystals could only be obtained from sitting drops containing equal volumes of 12–15  $\mu$ l protein and well solution under the optimized conditions of 6.5–7.5% PEG 3500, 17–18% isopropanol and 0.1 M sodium citrate (pH 5.6), with the addition of 4–5% glycerol to slow crystal growth and improve crystal size and shape. Maximum single-crystal dimensions reached 450  $\mu$ m  $\times$  150  $\mu$ m  $\times$  40  $\mu$ m. Before cooling the crystals to 100 K in liquid nitrogen, three rounds of increases in PEG 3350 concentration (12 h for each increase of 8% per cycle) were carried out in the mother liquor<sup>36</sup>. The final PEG 3350 concentration of about 31% was sufficient for cryoprotection. Crystals are summarized in Supplementary Table 1. There are two complexes per asymmetric unit, with a Matthews coefficient of 2.9 Å<sup>3</sup> Da<sup>-1</sup>, giving a solvent content of 57.8%.

To prepare Se-Met proTGF- $\beta$ 1, cells were washed with PBS (pH 7.4), supplemented with 1% FBS, then incubated for 8 h with methionine-free  $\alpha$ -MEM (SAFC Biosciences) supplemented with 50 mg l<sup>-1</sup> L-Se-Met (Sigma) and 10% dialyzed FBS. After replacement with the same medium, cells were cultured for 4 d. Se-Met proTGF- $\beta$ 1, at a yield of 1.5 mg l<sup>-1</sup>, was purified and crystallized identically to native proTGF- $\beta$ 1. Furthermore, a heavy-atom derivative was obtained by soaking crystals in mother liquor containing 0.4 mM HgBr<sub>2</sub> for 4 h.

**Structure determination and refinement.** Native Se multiple-wavelength anomalous dispersion (MAD) and single-wavelength anomalous dispersion (SAD) Hg derivative data were collected at 100 K at beamline 23-ID, then processed using HKL2000 (ref. 37) and XDS<sup>38</sup>. Statistics are in Supplementary Table 1. Initial experimental phases were determined independently using Se-MAD and Hg-SAD, with 19 out of 24 Se sites and 14 Hg sites in the asymmetric unit located using PHENIX<sup>39</sup>. Electron density maps from Se-Met phasing, calculated after fourfold non-crystallographic symmetry (NCS) averaging, clearly defined the orientation of each monomer. The mature TGF- $\beta$ 1 homodimer was easily docked into the map using model 1KLC with MOLREP<sup>40</sup> in CCP4 (ref. 41). The prodomain was built into the map manually. A crude model of proTGF- $\beta$ 1 was obtained after rigid-body refinement by PHENIX, with both domains as one rigid body. The same model was docked into Hg-SAD density for the two homodimers, using MOLREP.

To improve the phases and extend them to higher resolution, multi-crystal averaging (two crystals in total: Se-MAD and Hg derivative), multi-domain averaging (with separate masks for each prodomain and TGF- $\beta$  monomer) and solvent flattening and histogram matching were performed using DMMULTI<sup>42</sup> from the CCP4 suite. The mask for each domain was calculated by NCSMASK in

CCP4, and NCS matrices for each domain between molecules and crystals were computed by LSQKAB in CCP4. Rigid-body refinements were carried out by PHENIX for each lattice, on the basis of the averaged maps. The new models for each lattice were then used to calculate a set of new NCS matrices for the next cycle of DMMULTI. These steps were cycled twice.

Model building in COOT<sup>43</sup> was based on multi-crystal, multi-domain averaged electron density maps and  $2F_o - F_c$  maps. NCS restraints and translation-libration-screw (TLS) groups were used in refinement with PHENIX. The sequence-to-structure register was confirmed using Se anomalous maps. The multi-crystal, multi-domain averaging was repeated using the refined structure at an  $R_{\text{free}}$  of about 33% and no major differences were found. Two residues from the 3C protease site remain at the N terminus after cleavage. The structures include residues 0–62, 70–208, 216–241, 250–361 and one N-acetylglucosamine (NAG) residue (chain A); residues 1–62, 70–208, 216–242, 250–299, 310–361 and two NAG residues (chain B); residues 1–62, 68–208, 216–241, 250–361 and three NAG residues (chain C) and residues 0–62, 69–208, 216–242, 250–361 and two NAG residues (chain D). Validation and Ramachandran statistics used MOLPROBITY<sup>44</sup>. All structure figures were generated using Pymol (DeLano Scientific).

In the asymmetric unit of the crystal, a second pro-TGF dimer extends each two-stranded  $\beta$ -ribbon to form a four-stranded, inter-dimer super- $\beta$ -sheet in which Leu 203 forms a hydrophobic lattice contact. In its absence, Leu 203 may mediate hydrophobic interactions within the bowtie.

**Mutagenesis.** Wild-type human proTGF- $\beta$ 1 was inserted into the pEF1-puro plasmid. Site-directed mutagenesis was performed using QuikChange (Stratagene). All mutations were confirmed by DNA sequencing.

**Free cysteine labelling and prodomain detection.** HEK-293T cells were transfected using Polyfect reagent (Qiagen) according to the manufacturer's instructions, using 2  $\mu$ g of proTGF- $\beta$ 1 cDNA per 6-cm dish of cells at 70–80% confluency. The cells were then cultured in FreeStyle serum-free medium (Invitrogen) for 3 d. Supernatant was reacted with 450  $\mu$ M biotin-BMCC (1-biotinamido-4-(4'-(maleimidoethyl)cyclohexane)-carboxamido)butane (Pierce) for 60 min at 22 °C, followed by the addition of 40 mM N-ethylmaleimide. ProTGF- $\beta$ 1 was immunoprecipitated with 1.5  $\mu$ g anti-human-prodomain-1 (LAP-1) antibody (R&D Systems) and Protein A Sepharose beads (GE Healthcare) at 4 °C for 2 h, then subjected to reducing SDS 10% PAGE. After transfer to polyvinylidene difluoride membranes (Millipore), biotin was detected using streptavidin-horseradish peroxidase with the ECL-plus western blotting kit (GE Healthcare). Total proTGF- $\beta$ 1 was similarly detected on a separate blot using biotinylated human prodomain-1 (LAP-1) antibody.

**TGF- $\beta$ 1 activation assay.** Transformed mink lung epithelial cells (TMLCs) stably transfected with a luciferase construct under plasminogen activator inhibitor promoter 1 (ref. 5) were provided by D. Rifkin (New York University). HEK-293T cell transfectants stably expressing  $\alpha_v$  and  $\beta_6$  were selected with puromycin and G418. Clones expressing high levels of integrin  $\alpha_v\beta_6$  were selected by immunofluorescent flow cytometry using an anti- $\beta_6$  antibody. Cells stably transfected with empty vector were used as a control. These cells were subsequently transiently transfected with human wild-type or mutant proTGF- $\beta$ 1, using lipofectamine with 0.4  $\mu$ g plasmid DNA per well in a 48-well plate. After 16–24 h, each well was used to seed 3 wells of a 96-well plate with about 15,000 cells, which were co-cultured with 15,000 TMLCs in 100  $\mu$ l DMEM with 0.1% BSA for 16–24 h. TGF- $\beta$ 1-induced luciferase activity in cell lysates was measured using the luciferase assay system (Promega). To assess heat-releasable TGF- $\beta$ 1, cells were transfected as above except that polyfect was used in 6-well plates. After 2 d, cells were collected and heated in 150  $\mu$ l of DMEM with 0.1% BSA at 80 °C for 10 min. TGF- $\beta$ 1 activity in 50  $\mu$ l aliquots was measured using the luciferase assay with TMLCs.

**Negative-stain electron microscopy.** A large latent-complex fragment was isolated from supernatants of 293T cells transiently co-transfected with native human proTGF- $\beta$  and a human LTBP fragment containing the same N-terminal tag as was used above on proTGF- $\beta$ . The LTBP fragment contained the TB3 and TB4 domains and two intervening EGF-like domains (residues Thr 1333–Asn 1578, immature numbering). Multi-angle light scattering gave an  $M_r$  of 119,400, compared to a calculated  $M_r$  of 120,400 for a 2:1 proTGF- $\beta$ :LTBP fragment (2  $\times$  46,400 for proTGF- $\beta$ , including 2  $\times$  7,500 for three high-mannose N-linked sites, plus 27,600 for the LTBP fragment). The  $\alpha_v\beta_6$  ectodomain was expressed using C-terminal  $\alpha$ -helical coiled-coils and tags; purified, then subjected to negative-stain electron microscopy as previously described<sup>45</sup>. Purified proTGF- $\beta$ 1 or proTGF- $\beta$ 1 in complex with an LTBP fragment (20  $\mu$ g), proTGF- $\beta$ 1 (30  $\mu$ g) in molar excess over clasped  $\alpha_v\beta_6$  (20  $\mu$ g), or clasped  $\alpha_v\beta_6$  (60  $\mu$ g) in excess over proTGF- $\beta$  (10  $\mu$ g) were subjected to Superdex S200 chromatography in TBS (pH 7.5) with 1 mM Ca<sup>2+</sup> and 1 mM Mg<sup>2+</sup>. Peak fractions corresponding to the purified proteins or complexes were subjected to

negative-stain electron microscopy. Particle selection, alignment, classification and averaging were conducted as previously described<sup>46</sup>.

32. Keefe, A. D., Wilson, D. S., Seelig, B. & Szostak, J. W. One-step purification of recombinant proteins using a nanomolar-affinity streptavidin-binding peptide, the SBP-Tag. *Protein Expr. Purif.* **23**, 440–446 (2001).
33. Zou, Z. & Sun, P. D. Overexpression of human transforming growth factor- $\beta$ 1 using a recombinant CHO cell expression system. *Protein Expr. Purif.* **37**, 265–272 (2004).
34. Gentry, L. E. *et al.* Type 1 transforming growth factor beta: amplified expression and secretion of mature and precursor polypeptides in Chinese hamster ovary cells. *Mol. Cell. Biol.* **7**, 3418–3427 (1987).
35. Brunner, A. M. *et al.* Site-directed mutagenesis of glycosylation sites in the transforming growth factor-beta 1 (TGF beta 1) and TGF beta 2 (414) precursors and of cysteine residues within mature TGF beta 1: effects on secretion and bioactivity. *Mol. Endocrinol.* **6**, 1691–1700 (1992).
36. Heras, B. & Martin, J. L. Post-crystallization treatments for improving diffraction quality of protein crystals. *Acta Crystallogr. D* **61**, 1173–1180 (2005).
37. Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326 (1997).
38. Kabsch, W. in *International Tables for Crystallography, Vol. F: Crystallography of Biological Macromolecules* (eds Rossmann, M. G. & Arnold, E. V.) Ch. 25.2.9 XDS, 730–734 (Springer, 2001).
39. Adams, P. D. *et al.* PHENIX: building new software for automated crystallographic structure determination. *Acta Crystallogr. D* **58**, 1948–1954 (2002).
40. Vagin, A. & Teplyakov, A. Molecular replacement with MOLREP. *Acta Crystallogr. D* **66**, 22–25 (2010).
41. Bailey, S. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D* **50**, 760–763 (1994).
42. Cowtan, K. Recent developments in classical density modification. *Acta Crystallogr. D* **66**, 470–478 (2010).
43. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **60**, 2126–2132 (2004).
44. Davis, I. W. *et al.* MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res.* **35**, W375–W383 (2007).
45. Takagi, J., Petre, B. M., Walz, T. & Springer, T. A. Global conformational rearrangements in integrin extracellular domains in outside-in and inside-out signaling. *Cell* **110**, 599–611 (2002).
46. Chen, X. *et al.* Requirement of open headpiece conformation for activation of leukocyte integrin  $\alpha_x\beta_2$ . *Proc. Natl Acad. Sci. USA* **107**, 14727–14732 (2010).

# X-ray structure of a bacterial oligosaccharyltransferase

Christian Lizak<sup>1,2</sup>, Sabina Gerber<sup>2</sup>, Shin Numao<sup>1†</sup>, Markus Aebi<sup>1</sup> & Kaspar P. Locher<sup>2</sup>

Asparagine-linked glycosylation is a post-translational modification of proteins containing the conserved sequence motif Asn-X-Ser/Thr. The attachment of oligosaccharides is implicated in diverse processes such as protein folding and quality control, organism development or host-pathogen interactions. The reaction is catalysed by oligosaccharyltransferase (OST), a membrane protein complex located in the endoplasmic reticulum. The central, catalytic enzyme of OST is the STT3 subunit, which has homologues in bacteria and archaea. Here we report the X-ray structure of a bacterial OST, the PglB protein of *Campylobacter lari*, in complex with an acceptor peptide. The structure defines the fold of STT3 proteins and provides insight into glycosylation sequon recognition and amide nitrogen activation, both of which are prerequisites for the formation of the N-glycosidic linkage. We also identified and validated catalytically important, acidic amino acid residues. Our results provide the molecular basis for understanding the mechanism of N-linked glycosylation.

It is estimated that more than half of all eukaryotic proteins are glycoproteins; that is, specific amino acid side chains are chemically modified with carbohydrates in a process termed glycosylation<sup>1,2</sup>. The most abundant of these modifications is asparagine-linked (N-linked) glycosylation, which affects a multitude of cellular functions<sup>3–5</sup>. Asparagines are specifically glycosylated in the context of a consensus sequon Asn-X-Ser/Thr when located in the endoplasmic reticulum (ER). The reaction takes place on the luminal surface of the ER membrane and is catalysed by OST, a hetero-oligomeric membrane protein complex in most eukaryotes<sup>6</sup>. A hallmark of N-linked glycosylation is its broad specificity with respect to the polypeptide substrate, which is a direct consequence of the short recognition sequon<sup>7</sup>. This distinguishes OST from O-glycosyltransferases that modify serine or threonine residues and exhibit a higher specificity for their protein substrates<sup>8</sup>.

The key step of OST-catalysed glycosylation is the formation of an N-glycosidic linkage between the amide nitrogen of the acceptor asparagine and the C1 carbon of the first saccharide moiety of a lipid-linked oligosaccharide (LLO) donor (Fig. 1a). This results in the en bloc transfer of the oligosaccharide onto the acceptor asparagine. Details of the reaction mechanism are poorly understood owing to the absence of structural insight into OST at high resolution, but also to the complex chemical nature of LLO, its low abundance in biological samples, and its insolubility in water. In contrast, crystal structures of various soluble O-glycosyltransferases have been published and their reaction mechanisms investigated in detail<sup>9–12</sup>. For OST, the currently accepted model suggests that glycosylation sequons are recognized when in unfolded protein segments<sup>13–15</sup>, either during protein translocation into the ER or after translocation is completed<sup>16</sup>. The central, catalytically active component within OST is the STT3 subunit<sup>6,17</sup>. The other subunits are thought to assist and refine the reaction<sup>18,19</sup>.

N-linked glycosylation is not restricted to eukaryotes. Homologous processes are found in archaea and in defined taxa of proteobacteria<sup>20,21</sup>. Eukaryotic and bacterial LLOs contain isoprenoid moieties that anchor the oligosaccharides in the membrane, and pyrophosphates as leaving groups of the substitution reactions. However, the

attached oligosaccharides are chemically distinct, and unlike their eukaryotic counterparts, the OSTs of prokaryotes (and of eukaryotic kinetoplastids) consist of a single subunit, which is homologous to the STT3 subunits of eukaryotic OST complexes (Supplementary Fig. 1)<sup>22–24</sup>. The best-studied bacterial OST, termed PglB (84 kDa), is encoded in the protein glycosylation locus *pgl* of the Gram-negative bacterium *Campylobacter jejuni*. This gene cluster is sufficient for catalysing protein glycosylation when transferred into *Escherichia coli* cells<sup>25</sup>. The similarity in sequence and membrane topology indicates that PglB and eukaryotic STT3s share a common reaction mechanism<sup>6,25–27</sup>. To understand the molecular basis of N-linked glycosylation, we have determined the X-ray structure of PglB from *C. lari*, which is 56% identical to that of *C. jejuni*<sup>28</sup>. *C. lari* PglB is functional when co-expressed with the *C. jejuni* *pgl* cluster in *E. coli* cells (Fig. 1b). We co-crystallized *C. lari* PglB with the hexapeptide DQNATF, an optimal acceptor sequence for *C. jejuni* PglB<sup>29</sup>. X-ray diffraction data was anisotropic and extended to 3.4 Å resolution. The structure was refined to *R*/*R*<sub>free</sub> values of 23.8% and 27.1%, respectively (Supplementary Table 1).

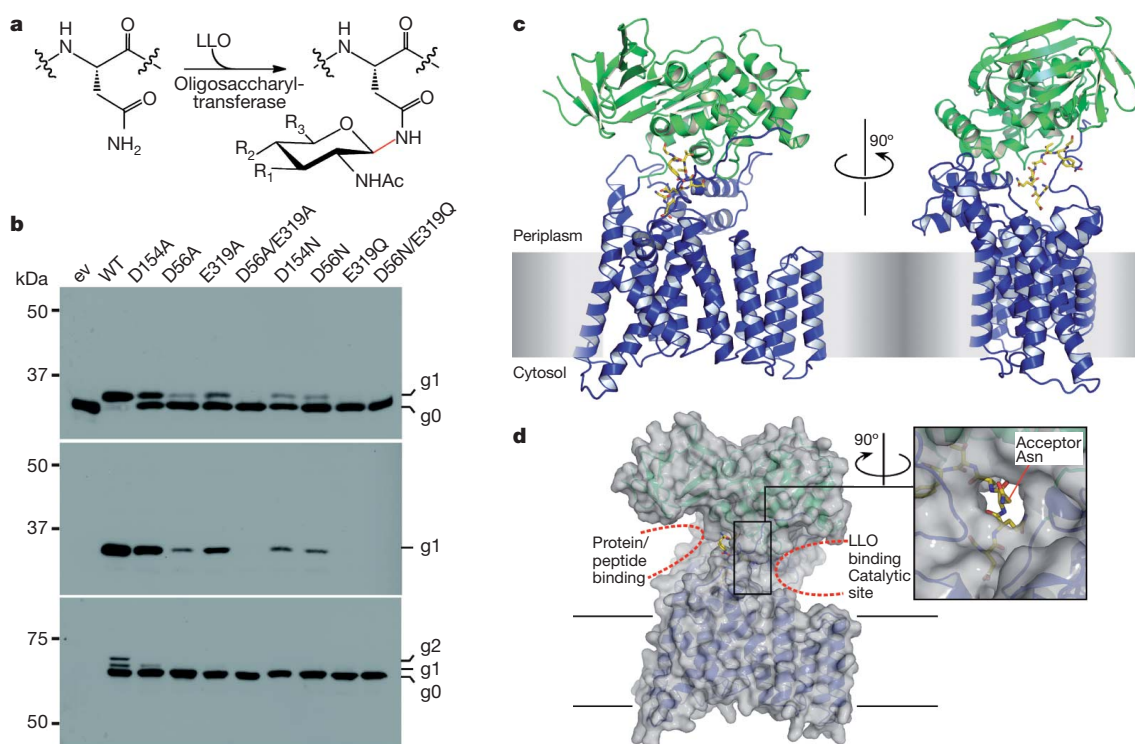
## Structure of *C. lari* PglB

In agreement with earlier predictions (reviewed in ref. 6), the structure of PglB revealed two domains: a transmembrane domain comprising residues 1–432 and a periplasmic domain comprising residues 433–712 (Fig. 1c). In addition to the covalent linkage, the two domains have extensive non-covalent interactions, provided mainly by the first external loop (EL1) of the transmembrane domain that forms two helices parallel to the membrane plane. The periplasmic domain features a mixed  $\alpha/\beta$  fold that was previously observed in the structures of the homologous domains of *C. jejuni* PglB and of *Pyrococcus furiosus* AglB<sup>30,31</sup>. However, these isolated domains were catalytically inactive and unable to bind acceptor peptide. Our structure of full-length PglB provides a molecular explanation by revealing that the transmembrane domain is indispensable both for peptide binding and catalysis.

In contrast to the periplasmic domain, the transmembrane domain features a novel fold (Fig. 2). Thirteen transmembrane segments are

<sup>1</sup>Institute of Microbiology, Department of Biology, ETH Zurich, CH-8093 Zurich, Switzerland. <sup>2</sup>Institute of Molecular Biology and Biophysics, Department of Biology, ETH Zurich, CH-8093 Zurich, Switzerland. †Present address: Novartis Pharma AG, CH-4002 Basel, Switzerland.





**Figure 1 | Activity and structure of *C. lari* PglB.** **a**, Reaction scheme of N-linked glycosylation, yielding an N-glycosidic bond (red). In bacteria,  $R_1$  = oligosaccharyl,  $R_2$  = NH-Ac,  $R_3$  =  $\text{CH}_3$ . In eukaryotes,  $R_1$  = OH,  $R_2$  = oligosaccharyl,  $R_3$  =  $\text{CH}_2\text{OH}$ . **b**, *In vivo* glycosylation assay in *E. coli*. Immunoblots detecting acceptor protein 3D5 (top), glycans (middle) or PglB (bottom). PglB constructs indicated above the lanes include vector control (ev), wild type (WT), or PglB mutants. Glycosylation yields a mobility shift from the unmodified (g0) to the glycosylated form (g1). Functional PglB is partially auto-

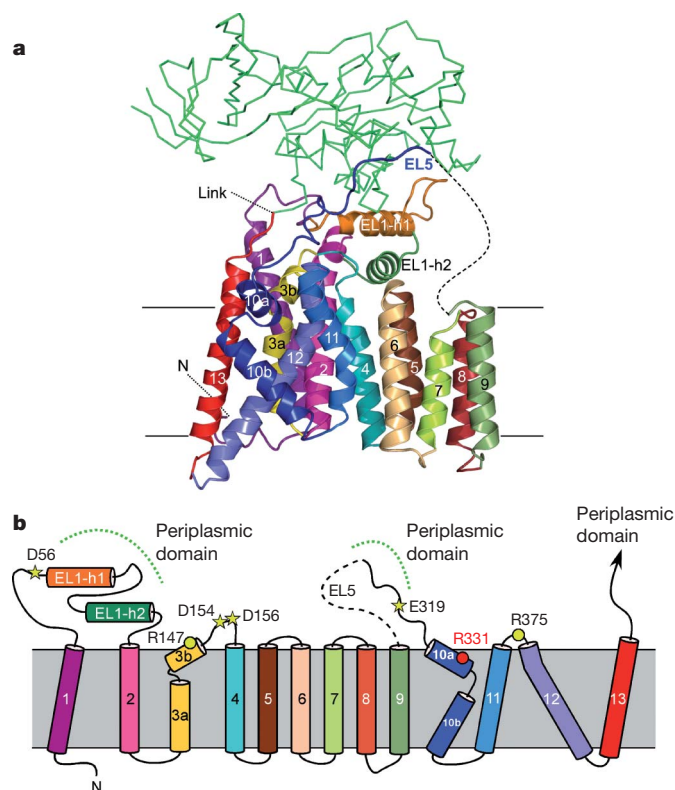
glycosylated at N535 and N556, resulting in two additional bands (g1 and g2). **c**, Ribbon diagrams of PglB structure, with transmembrane and periplasmic domains in blue and green, respectively, and bound acceptor peptide in stick representation and coloured yellow. The presumed position of the membrane is indicated by a grey rectangle. **d**, Surface representation in semi-transparent grey, with ribbons as in **a**. Two cavities at opposite sides of PglB are indicated by dashed lines, providing access for substrates. The cavities are connected by a tunnel that accommodates the acceptor asparagine.

connected by short cytoplasmic and external loops, with the exception of the long external loops EL1 and EL5. Whereas EL1 is well ordered, EL5 is only partially ordered, with 25 residues disordered in the electron density map. Transmembrane segments TM1–4 and TM10–13 form the sequon-binding and catalytic sites and provide the bulk of the interface with the periplasmic domain. In the peptide-bound state, PglB forms two large cavities above the membrane surface, located at opposite sides of the protein (Fig. 1d). The left-side cavity provides access for acceptor proteins as suggested by the presence of bound peptide in the structure, whereas the right-side cavity harbours the catalytic residues and probably serves as the binding pocket for LLO. The two cavities are connected where the side chain of the acceptor asparagine reaches from the peptide-binding site into the catalytic pocket.

### Acceptor sequon binding and recognition

Clear density for the bound hexapeptide DQNATF was observed in a location that placed the acceptor asparagine some 15 Å above the membrane surface (Fig. 1c). Almost 80% of the contact surface of the peptide (calculated by areaimol<sup>32</sup>) is buried at the interface of the transmembrane and periplasmic domains of PglB (Fig. 3a), indicating tight binding and a firmly imposed conformation. The peptide forms a loop that almost completes a 180° turn; accordingly, protein substrates have to present their glycosylation sequons in sufficiently large, flexible and surface-exposed loops<sup>33</sup>, because the peptide-binding cavity of PglB does not appear to fit fully folded protein domains. The observed conformation of the peptide would be incompatible with a proline residue at the +1 position, in agreement with the observation that +1 prolines are not allowed in glycosylation sequons<sup>34,35</sup>.

A hallmark of N-linked glycosylation is the requirement of a serine or threonine at the +2 position of the acceptor sequon. The PglB structure provides a molecular explanation by revealing that the  $\beta$ -hydroxyl group of the +2 Thr of bound hexapeptide forms three hydrogen bonds, one with each of the side chains of the 'WWD motif', which is strictly conserved in STT3 proteins (Fig. 3b and Supplementary Fig. 3). The motif is located in the periplasmic domain, and the interaction of the two tryptophan and the aspartate side chains saturate the hydrogen-bonding capacity of the  $\beta$ -hydroxyl group. The arrangement physically separates the +2 Thr from the acceptor asparagine, and we conclude that the WWD motif defines the polypeptide substrate specificity, but is not directly involved in catalysis. Notably, the structure can also explain preferences and deviations at the +2 position of glycosylation sequons. The  $\gamma$ -methyl group of the +2 Thr is in van der Waals contact with Ile 572 of PglB (3.6 Å distance to the  $\gamma$ -methyl group of Ile 572; Fig. 3b and Supplementary Fig. 4). This stabilizing interaction is absent if a serine is in the +2 position, which may explain that acceptor sequons containing a +2 Thr are glycosylated 40 times more efficiently than if they contain a +2 Ser (ref. 36). The structure further suggests that the non-natural, S-configured threonine would cause a steric clash with Ile 572. S-configured threonine is indeed not allowed at the +2 position, with a 15,000-fold reduction in glycosylation efficiency compared to R-configured threonine<sup>37</sup>. Ile 572 is conserved in bacteria and has been suggested to be part of a MXXI motif<sup>31</sup>. However, the corresponding residue in the archaeal AglB protein was found to be a lysine<sup>30</sup>, and sequence alignments with eukaryotic STT3 homologues reveal no clear conservation of Ile 572, indicating that residues other than isoleucine can provide contacts to the +2 Thr in homologous proteins. The PglB structure can also explain allowed deviations from



**Figure 2 | Topology of transmembrane domain.** **a**, Ribbon diagram, with helices numbered and coloured as in **b**. The periplasmic domain is shown as a green backbone trace. **b**, Topological schematic indicating helices and connecting loops. Dashed green lines indicate non-covalent contacts to the periplasmic domain. Conserved residues forming the active site of PglB are indicated by yellow stars and spheres and are labelled. R331 contributing to peptide recognition is indicated in red.

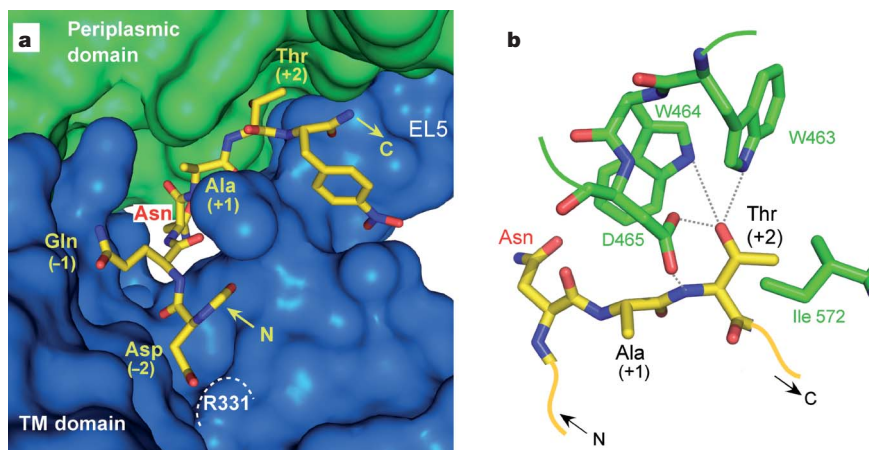
the consensus sequon: the acceptor sequence N-X-C, present in ~2.2% of experimentally determined glycosylation sites of the mouse glycoproteome<sup>2</sup>, is probably allowed because the  $\beta$ -sulphhydryl group of cysteine can form similar hydrogen bonds as a  $\beta$ -hydroxyl group. Glycines, alanines and valines have also been reported at the +2 position of glycosylated sequons, albeit only at low abundance<sup>2,28,38</sup>. These residues can, in principle, be accommodated in the binding pocket of PglB because they are equal in size or smaller than

threonine. However, glycosylation of sequons such as N-G-X, with X being larger than threonine, or of T/S-X-N (reverse sequons)<sup>2</sup> cannot be explained by the PglB structure.

Compared to the eukaryotic enzymes, bacterial OSTs analysed thus far have an extended acceptor sequon: glycosylation is only efficient if a negatively charged residue (Asp or Glu) is present at the -2 position, resulting in a consensus sequon D/E-X-N-X-S/T<sup>35</sup>. In PglB, R331 provides a salt bridge to the -2 Asp of bound peptide (Fig. 3a and Supplementary Fig. 4), thereby strengthening the PglB-peptide interaction. R331 is conserved in bacteria, but not in eukaryotes, where no requirement for a negative charge at the -2 position is observed. The extended sequon recognition may reflect the need for tighter peptide binding in bacteria, where the local concentration of the acceptor polypeptide is probably lower than in eukaryotes.

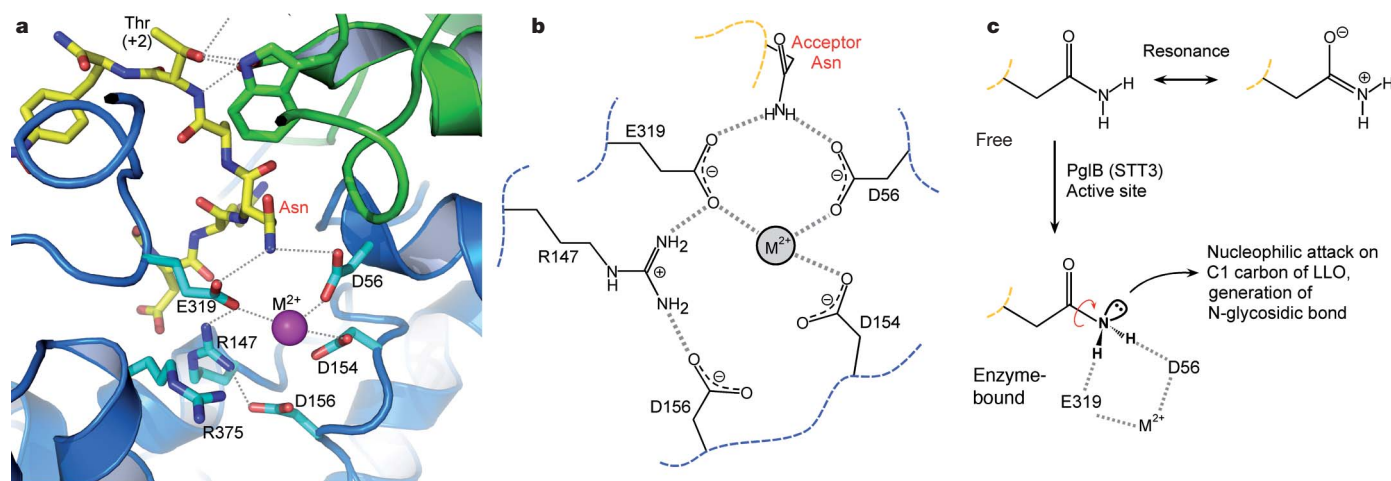
### Catalytic site site

The catalytic pocket is located in the right-side cavity of PglB (Fig. 1d) and is marked by a bound cation, located ~8 Å above the membrane boundary. Owing to the high concentration of magnesium salt in the crystallization solution, it was modelled as  $Mg^{2+}$ . PglB, like all OSTs, is only functional with a bound divalent cation ( $M^{2+}$ )<sup>39,40</sup>. The physiological cation was suggested to be  $Mn^{2+}$ , but PglB is also active in  $Mg^{2+}$ , a property that was previously observed for other metal-dependent glycosyltransferases<sup>11</sup>. The catalytic pocket of PglB features three acidic side chains (D56, D154, E319) that are part of the transmembrane domain and coordinate  $M^{2+}$  (Fig. 4 and Supplementary Fig. 5). At the current resolution, water molecules that might be additional ligands of  $M^{2+}$  cannot be modelled. The residues located in the catalytic pocket are generally conserved in STT3 proteins (Supplementary Fig. 3). The aspartates D154 and D156 belong to a previously reported D-X-D motif, and mutation of either aspartate to an alanine abolished the activity of the mannosyltransferase GPI-MT-1, a member of the same glycosyltransferase superfamily as PglB (GT-C)<sup>41,42</sup>. In contrast, D56 and E319 have not been previously identified as catalytically relevant, but their carboxyl groups interact both with  $M^{2+}$  and the amido group of the acceptor asparagine. Such interactions would not be possible if a glutamine or an aspartate side chain was present instead of the acceptor asparagine, which may explain why glutamines and aspartates are not glycosylated by OST. To confirm the catalytic involvement of the three  $M^{2+}$ -binding residues, we mutated them individually to alanines and tested the activity of the resulting PglB mutants in a complementation assay (Fig. 1b). Even though OST activity is not limiting in our assay, the mutation D154A reduced the observed glycosylation yield by >50%, whereas D56A



**Figure 3 | Sequon binding and recognition.** **a**, Transmembrane (TM) and periplasmic domains of PglB are in blue and green, respectively. Acceptor peptide is in ball and stick representation; N and C denote amino and carboxy termini. Peptide residues are labelled yellow, acceptor Asn is labelled red. **b**, PglB residues interacting with the +2 Thr of bound peptide are shown in green and labelled. Hydrogen bonds from the WWD motif to the  $\beta$ -hydroxyl group are indicated by dashed lines.





**Figure 4 | Catalytic site and amide nitrogen activation.** **a**, Transmembrane and periplasmic domains of PglB are coloured blue and green, respectively. Selected side chains are in ball and stick representation, with carbon atoms coloured cyan and green for transmembrane and periplasmic domain residues, respectively, and yellow for acceptor peptide. Grey dashed lines indicate hydrogen bonds or interactions with the divalent cation  $M^{2+}$ . **b**, Chemical structure of the catalytic site, indicating interactions as in **a**. Blue and yellow

dashed lines indicate transmembrane domain and peptide backbones, respectively. **c**, Presumed mechanism of amide nitrogen activation. Yellow dashed lines indicate peptide backbone. The amido group of free acceptor Asn features electron delocalization, indicated by resonance. When bound to PglB, the amido group may form hydrogen bonds with the catalytically essential D56 and E319, requiring rotation around the C–N bond (red arrow).

and E319A reduced it by >90%. A PglB double mutant D56A/E319A was completely inactive.

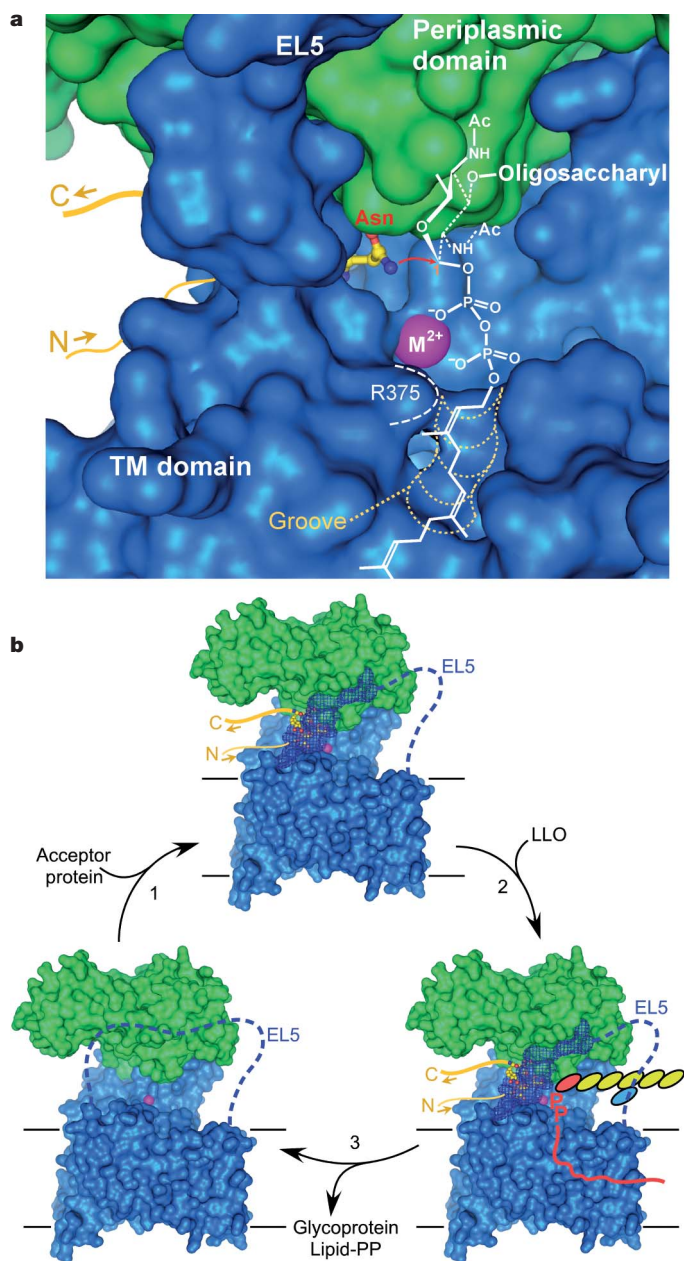
There is a controversial discussion on how the amido group of the acceptor asparagine might be activated for a nucleophilic attack on the C1 carbon of LLO substrate. Amides are poor nucleophiles because the free electron pair of the nitrogen is conjugated to the carbonyl group. As a consequence, the N–C bond has double-bond character, and the nucleophilicity of the nitrogen is low. To explain the reactivity of the amido group, specific conformations of the acceptor peptide, such as a ‘ $\beta$ -turn’ or an ‘Asx turn’, have been proposed, invoking direct involvement of the +2 Ser/Thr in catalysis<sup>34,43</sup>. Given the firm binding of the +2 Thr to the WWD motif in our PglB structure, we can rule out such an involvement. Instead, the structure of PglB presents a distinct possibility for explaining amide nitrogen activation: the catalytically essential residues D56 and E319 are optimally positioned to form hydrogen bonds with the two amide protons of the acceptor asparagine. Forming such hydrogen bonds would require a rotation of the N–C bond of the amido group, thereby abolishing the conjugation of the nitrogen electrons with the carbonyl group (Fig. 4c). Not only would this increase the electronegative nature of the amide nitrogen (by polarizing the N–H bonds and increasing the electron density on the nitrogen), but it would also generate an  $sp^3$  hybridized nitrogen with a reactive lone pair optimally positioned for a nucleophilic attack. The energy barrier for rotating the N–C bond in most amides is estimated to be 16–20 kcal mol<sup>−1</sup>, and the 270° amide conformation shown in Fig. 4c has been calculated to have an energy of ~18.6 kcal mol<sup>−1</sup> relative to the planar conformation<sup>44</sup>. Hence, it would take 1–2 low-barrier hydrogen bonds<sup>45</sup> (each worth ~10 kcal mol<sup>−1</sup>) to provide sufficient energy to break permanently the conjugation of the carboxamide group of the acceptor asparagine. The carboxyl groups of D56 or E319 might provide such interactions in the transition state of the reaction. However, it will require a higher resolution structure to measure reliably the lengths of the hydrogen bonds. Mutating D56 to asparagine (D56N) has an even more pronounced inhibitory effect than truncation to alanine, and the E319Q mutant is completely inactive (Fig. 1b). This demonstrates that the negative charges provided by the carboxyl groups of D56 and E319 are essential for catalysis and that the acidic side chains cannot be replaced by iso-electronic amides. Steric effects might explain the increased inhibition of D56N relative to D56A and of E319Q relative to E319A.

## Glycosylation mechanism

Given that PglB is active even when solubilized in detergent (used for crystallization), our structure has probably captured a functionally competent state. Glycosylation occurs with inversion of configuration at the substituted C1 carbon. We modelled the LLO substrate into the PglB structure such that the di-*N*-acetyl-bacillosamine moiety is properly aligned for a nucleophilic attack by the activated amide nitrogen, while the leaving pyrophosphate group is in contact with  $M^{2+}$  and the conserved R375 (Fig. 5a and Supplementary Fig. 3). This arrangement places the additional saccharide moieties in the right-hand cavity of PglB and the C2 substituent of the first saccharide moiety, a *N*-acetyl group present in bacterial and eukaryotic LLO, in the vicinity of a conserved tyrosine residue (Y468, Supplementary Fig. 3), where density consistent with a bound water molecule is observed. When modelled as shown in Fig. 5a, the isoprenoid moieties of LLO are embedded in a hydrophobic groove on the PglB surface, pointing into the lipid bilayer (Supplementary Fig. 6). The function of  $M^{2+}$  in PglB appears to be twofold: on the one hand, it orients the acidic residues D56 and E319 that interact with the acceptor asparagine, and on the other, it stabilizes the leaving group (lipid-pyrophosphate) of the substitution. The dual function seems to be distinct from metal-dependent, configuration-inverting glycosyltransferases of the GT-A family, where  $M^{2+}$  primarily serves the stabilization of the leaving group<sup>11</sup>.

With the acceptor peptide present in the structure and the LLO molecule tentatively modelled, we can propose a basic, three-state catalytic cycle for PglB-catalysed glycosylation (Fig. 5b). A critical element of the proposed mechanism is the engagement and disengagement of EL5, which we expect to be flexible and disordered in the absence of bound acceptor peptide (ground state). Upon sequon binding, the C-terminal half of EL5 is ordered and pins the bound acceptor peptide against the periplasmic domain of PglB, thereby restricting its motion. Because the essential E319 is part of EL5, this simultaneously results in the formation of the catalytic site, where the acceptor asparagine is oriented and the amide nitrogen activated. Binding of LLO will then result in a nucleophilic attack of the activated amide nitrogen, resulting in glycosylation. Once the glycosidic bond is formed, the newly attached saccharides are tightly pressed against PglB, causing steric tension that can be released by disengagement of EL5. This allows the glycopeptide to dissociate from the enzyme. Subsequent cleavage of the lipid-linked pyrophosphate





**Figure 5 | Proposed glycosylation mechanism of PglB.** Transmembrane (TM) and periplasmic domain surfaces are shown in blue and green, respectively. Bound acceptor peptide is in ball and stick representation, and yellow lines indicate the N and C termini. **a**, The chemical structure of bacterial LLO is shown in white, highlighting presumed interactions with  $M^{2+}$  and R375 of PglB, and a collinear arrangement of attacking and leaving groups of the substitution. A red arrow indicates the nucleophilic attack. A predominantly hydrophobic groove accommodating the isoprenoid moieties is indicated. **b**, The observed crystal structure reflects the top state, with acceptor peptide bound and the C-terminal half of EL5 ordered and engaged. The bottom-left state reflects the ground state, with no substrates bound and EL5 disordered, as indicated by a dashed line. In the bottom-right state, bound *C. jejuni* LLO is shown (red line for isoprenoid moieties, P for phosphate, ellipsoids for saccharide moieties). The molecular events indicated by arrows are: 1, sequon binding, EL5 engagement, acceptor asparagine activation; 2, LLO binding, glycosylation; 3, disengagement of EL5, release of glycosylated sequon and lipid-pyrophosphate (PP).

anhydride and folding of the newly glycosylated protein probably provide the main contributions to the driving force of the *in vivo* reaction. We should point out that there is no experimental evidence indicating a strict sequence of events. Thus, the binding of LLO to PglB might precede that of acceptor peptide.

## Conclusions and outlook

Our results provide the basis for understanding the mechanism of N-linked glycosylation at a molecular level. Future studies will be directed at probing the outlined chemical concepts, which are at the core of both bacterial and eukaryotic OST. In addition, the interaction of the catalytic STT3 with the additional subunits in eukaryotic OST needs to be defined. At present, the highest resolution structure of a eukaryotic OST is a 12 Å electron microscopy map of the yeast enzyme<sup>46</sup>. The PglB structure also provides opportunities for engineering the substrate specificity both with respect to the acceptor sequon and LLO, which may open up new avenues for the production of glycoprotein and glycoconjugate therapeutics.

## METHODS SUMMARY

**In vivo glycosylation assay.** *Escherichia coli* SCM6 cells were transformed with three separate plasmids containing: (1) the *C. jejuni* *pglB<sub>mut</sub>* cluster (containing an inactivated *pglB* gene) to generate and flip LLO; (2) the glycosylation acceptor protein 3D5, a single-chain Fv fragment containing a DQNAT acceptor sequon<sup>47</sup>; (3) *C. lari* PglB, wild type or mutants. Expression and glycosylation of 3D5 was monitored by SDS-PAGE of periplasmic cell extracts and visualized by mobility shift due to increased size in an immunoblot using anti-c-Myc antibody, or the reactivity of the glycoprotein in an anti-glycan immunoblot using hR6 antiserum. Expression of *C. lari* PglB was monitored by immunoblots of whole-cell extracts using anti-haemagglutinin (HA) antiserum.

**Crystallization and structure determination.** *Campylobacter lari* PglB containing a C-terminal His<sub>10</sub> tag was expressed in *E. coli* cells and purified in β-D-dodecyl maltoside. PglB was co-crystallized with an acceptor peptide Ac-DQNATF {4-NO<sub>2</sub>}-NH<sub>2</sub>, containing an acetylated N terminus and an amidated C terminus. For experimental phasing, crystals were soaked in ethyl mercury phosphate (EMP) before data collection. Three distinct data sets of EMP derivatives were collected. The phase problem was solved by a combination of MIRAS and molecular replacement, using the periplasmic domain of *C. jejuni* PglB (Protein Data Bank code 3AAG) as a search model.

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

Received 27 January; accepted 26 April 2011.

- Apweiler, R., Hermjakob, H. & Sharon, N. On the frequency of protein glycosylation, as deduced from analysis of the SWISS-PROT database. *Biochim. Biophys. Acta* **1473**, 4–8 (1999).
- Zielinska, D. F., Gnad, F., Wisniewski, J. R. & Mann, M. Precision mapping of an *in vivo* N-glycoproteome reveals rigid topological and sequence constraints. *Cell* **141**, 897–907 (2010).
- Helenius, A. & Aebi, M. Roles of N-linked glycans in the endoplasmic reticulum. *Annu. Rev. Biochem.* **73**, 1019–1049 (2004).
- Kornfeld, R. & Kornfeld, S. Assembly of asparagine-linked oligosaccharides. *Annu. Rev. Biochem.* **54**, 631–664 (1985).
- Varki, A. Biological roles of oligosaccharides—all of the theories are correct. *Glycobiology* **3**, 97–130 (1993).
- Kelleher, D. J. & Gilmore, R. An evolving view of the eukaryotic oligosaccharyltransferase. *Glycobiology* **16**, 47R–62R (2006).
- Petrescu, A. J., Milac, A. L., Petrescu, S. M., Dwek, R. A. & Wormald, M. R. Statistical analysis of the protein environment of N-glycosylation sites: implications for occupancy, structure, and folding. *Glycobiology* **14**, 103–114 (2004).
- Breton, C. & Imbert, A. Structure/function studies of glycosyltransferases. *Curr. Opin. Struct. Biol.* **9**, 563–571 (1999).
- Charnock, S. J. & Davies, G. J. Structure of the nucleotide-diphospho-sugar transferase, SpsA from *Bacillus subtilis*, in native and nucleotide-complexed forms. *Biochemistry* **38**, 6380–6385 (1999).
- Ünlilil, U. M. & Rini, J. M. Glycosyltransferase structure and mechanism. *Curr. Opin. Struct. Biol.* **10**, 510–517 (2000).
- Lairson, L. L., Henriessat, B., Davies, G. J. & Withers, S. G. Glycosyltransferases: Structures, functions, and mechanisms. *Annu. Rev. Biochem.* **77**, 521–555 (2008).
- Lazarus, M. B., Nam, Y., Jiang, J., Sliz, P. & Walker, S. Structure of human O-GlcNAc transferase and its complex with a peptide substrate. *Nature* **469**, 564–567 (2011).
- Chen, W. & Helenius, A. Role of ribosome and translocon complex during folding of influenza hemagglutinin in the endoplasmic reticulum of living cells. *Mol. Biol. Cell* **11**, 765–772 (2000).
- Silberstein, S. & Gilmore, R. Biochemistry, molecular biology, and genetics of the oligosaccharyltransferase. *FASEB J.* **10**, 849–858 (1996).
- Whitley, P., Nilsson, I. & von Heijne, G. A nascent secretory protein may traverse the ribosome endoplasmic reticulum translocase complex as an extended chain. *J. Biol. Chem.* **271**, 6241–6244 (1996).

16. Ruiz-Canada, C., Kelleher, D. J. & Gilmore, R. Cotranslational and Posttranslational N-glycosylation of polypeptides by distinct mammalian OST isoforms. *Cell* **136**, 272–283 (2009).
17. Zufferey, R. *et al.* Stt3, a highly conserved protein required for yeast oligosaccharyl transferase activity *in vivo*. *EMBO J.* **14**, 4949–4960 (1995).
18. Wilson, C. M. & High, S. Ribophorin I acts as a substrate-specific facilitator of N-glycosylation. *J. Cell Sci.* **120**, 648–657 (2007).
19. Schulz, B. L. *et al.* Oxidoreductase activity of oligosaccharyltransferase subunits Ost3p and Ost6p defines site-specific glycosylation efficiency. *Proc. Natl Acad. Sci. USA* **106**, 11061–11066 (2009).
20. Mescher, M. F. & Strominger, J. L. Purification and characterization of a prokaryotic glycoprotein from cell-envelope of *Halobacterium salinarum*. *J. Biol. Chem.* **251**, 2005–2014 (1976).
21. Szymanski, C. M., Yao, R. J., Ewing, C. P., Trust, T. J. & Guerry, P. Evidence for a system of general protein glycosylation in *Campylobacter jejuni*. *Mol. Microbiol.* **32**, 1022–1030 (1999).
22. Izquierdo, L. *et al.* Distinct donor and acceptor specificities of *Trypanosoma brucei* oligosaccharyltransferases. *EMBO J.* **28**, 2650–2661 (2009).
23. Nasab, F. P., Schulz, B. L., Gamarro, F., Parodi, A. J. & Aebi, M. All in one: *Leishmania major* STT3 proteins substitute for the whole oligosaccharyltransferase complex in *Saccharomyces cerevisiae*. *Mol. Biol. Cell* **19**, 3758–3768 (2008).
24. Szymanski, C. M. & Wren, B. W. Protein glycosylation in bacterial mucosal pathogens. *Nature Rev. Microbiol.* **3**, 225–237 (2005).
25. Wacker, M. *et al.* N-linked glycosylation in *Campylobacter jejuni* and its functional transfer into *E. coli*. *Science* **298**, 1790–1793 (2002).
26. Abu-Qarn, M. & Eichler, J. Protein N-glycosylation in Archaea: defining *Haloferax volcanii* genes involved in S-layer glycoprotein glycosylation. *Mol. Microbiol.* **61**, 511–525 (2006).
27. Li, L. *et al.* Overexpression and topology of bacterial oligosaccharyltransferase PglB. *Biochem. Biophys. Res. Commun.* **394**, 1069–1074 (2010).
28. Schwarz, F. *et al.* Relaxed acceptor site specificity of bacterial oligosaccharyltransferase *in vivo*. *Glycobiology* **21**, 45–54 (2011).
29. Chen, M. M., Glover, K. J. & Imperiali, B. From peptide to protein: Comparative analysis of the substrate specificity of N-linked glycosylation in *C. jejuni*. *Biochemistry* **46**, 5579–5585 (2007).
30. Igura, M. *et al.* Structure-guided identification of a new catalytic motif of oligosaccharyltransferase. *EMBO J.* **27**, 234–243 (2008).
31. Maita, N., Nyirenda, J., Igura, M., Kamishikiryo, J. & Kohda, D. Comparative structural biology of eubacterial and archaeal oligosaccharyltransferases. *J. Biol. Chem.* **285**, 4941–4950 (2010).
32. Collaborative Computational Project, Number 4. The CCP4 Suite: programs for protein crystallography. *Acta Crystallogr. D* **50**, 760–763 (1994).
33. Kowarik, M. *et al.* N-linked glycosylation of folded proteins by the bacterial oligosaccharyltransferase. *Science* **314**, 1148–1150 (2006).
34. Bause, E. & Legler, G. The role of the hydroxy amino acid in the triplet sequence Asn-Xaa-Thr(Ser) for the N-glycosylation step during glycoprotein biosynthesis. *Biochem. J.* **195**, 639–644 (1981).
35. Kowarik, M. *et al.* Definition of the bacterial N-glycosylation site consensus sequence. *EMBO J.* **25**, 1957–1966 (2006).
36. Bause, E. Model studies on N-glycosylation of proteins. *Biochem. Soc. Trans.* **12**, 514–517 (1984).
37. Breuer, W., Klein, R. A., Hardt, B., Bartoschek, A. & Bause, E. Oligosaccharyltransferase is highly specific for the hydroxy amino acid in Asn-Xaa-Thr/Ser. *FEBS Lett.* **501**, 106–110 (2001).
38. Valliere-Douglass, J. F. *et al.* Asparagine-linked oligosaccharides present on a non-consensus amino acid sequence in the C<sub>H1</sub> domain of human antibodies. *J. Biol. Chem.* **284**, 32493–32506 (2009).
39. Imperiali, B. & Rickert, K. W. Conformational implications of asparagine-linked glycosylation. *Proc. Natl Acad. Sci. USA* **92**, 97–101 (1995).
40. Sharma, C., Lehle, L. & Tanner, W. N-glycosylation of yeast proteins—characterization of the solubilized oligosaccharyl transferase. *Eur. J. Biochem.* **116**, 101–108 (1981).
41. Liu, J. & Mushegian, A. Three monophyletic superfamilies account for the majority of the known glycosyltransferases. *Protein Sci.* **12**, 1418–1431 (2003).
42. Maeda, Y. *et al.* PIG-M transfers the first mannose to glycosylphosphatidylinositol on the luminal side of the ER. *EMBO J.* **20**, 250–261 (2001).
43. Imperiali, B., Shannon, K. L. & Rickert, K. W. Role of peptide conformation in asparagine-linked glycosylation. *J. Am. Chem. Soc.* **114**, 7942–7944 (1992).
44. Wiberg, K. B. & Breneman, C. M. Resonance interactions in acyclic systems. 3. Formamide internal rotation revisited. Charge and energy redistribution along the C–N bond rotational pathway. *J. Am. Chem. Soc.* **114**, 831–840 (1992).
45. Cleland, W. W. & Kreevoy, M. M. Low-barrier hydrogen bonds and enzymatic catalysis. *Science* **264**, 1887–1890 (1994).
46. Li, H., Chavan, M., Schindelin, H., Lennarz, W. J. & Li, H. L. Structure of the oligosaccharyl transferase complex at 12 Å resolution. *Structure* **16**, 432–440 (2008).
47. Lizak, C., Fan, Y. Y., Weber, T. C. & Aebi, M. N-linked glycosylation of antibody fragments in *Escherichia coli*. *Bioconjug. Chem.* **22**, 488–496 (2011).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank the beamline staff at the Swiss Light Source for assistance with data collection, S. Fleurkens and M. Bucher for technical assistance, and D. Arigoni for discussions. *E. coli* SCM6 was provided by C. Marolda and M. Valvano. C.L. was affiliated with the Life Science Zurich Graduate School. This research was supported by the NCCR Structural Biology Zurich (grant to K.P.L.) and Swiss National Science Foundation grants to M.A. (SNF 31003A-127098/1) and K.P.L. (SNF 31003A-131075/1).

**Author Contributions** C.L., M.A. and K.P.L. designed the experiments. C.L., S.G. and S.N. performed the experiments; K.P.L. performed crystallographic calculations and model building; C.L., M.A. and K.P.L. analysed the data and wrote the manuscript.

**Author Information** Atomic coordinates and structure factors have been deposited with the Protein Data Bank under accession code 3RCE. Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare competing financial interests: details accompany the full-text HTML version of the paper at [www.nature.com/nature](http://www.nature.com/nature). Readers are welcome to comment on the online version of this article at [www.nature.com/nature](http://www.nature.com/nature). Correspondence and requests for materials should be addressed to K.P.L. ([locher@mol.biol.ethz.ch](mailto:locher@mol.biol.ethz.ch)).

## METHODS

**In vivo complementation.** To analyse the activity of PglB from *C. lari* in vivo the gene encoding PglB was amplified from the *pgl* gene cluster of *Campylobacter lari* isolate (sample provided by H. Hächler) genomic DNA by polymerase chain reaction (PCR) and was cloned into a pMLBAD plasmid<sup>48</sup> with a C-terminal HA tag fused to PglB, resulting in the plasmid pMIK71 (ref. 28). For complementation studies, pMIK71 or pMLBAD empty vector were transformed into *E. coli* SCM6 cells carrying the plasmids pCL21 (ref. 47) and pACYC*pgl*<sub>mut</sub> (ref. 25). pCL21 encodes for the expression of the single-chain Fv fragment of 3D5 carrying a DQNAT glycosylation site in the linker region and a C-terminal Myc tag fused to 3D5. pACYC*pgl*<sub>mut</sub> codes for the biosynthesis of the *C. jejuni* lipid-linked oligosaccharide (LLO) with an inactivated *C. jejuni* *pglB* gene (W458A and D459A). A 5 ml pre-culture was inoculated from a single clone and grown overnight at 37 °C in LB medium. The main culture was inoculated to an optical density ( $A_{600}$ ) of 0.05 in 15 ml LB medium and grown at 37 °C to  $A_{600}$  of 0.5. The culture was induced by addition of arabinose to 0.1% (w/v) and grown for 4 h at 24 °C. For extraction of periplasmic proteins, an equivalent of 1 ml culture volume with an  $A_{600}$  of 3 was harvested by centrifugation, re-suspended in 150 µl extraction buffer, consisting of 30 mM Tris-HCl, pH 8.5; 20% (w/v) sucrose; 1 mM EDTA and 1 mg ml<sup>-1</sup> lysozyme (Sigma) and incubated for 1 h at 4 °C. A final centrifugation step yielded periplasmic proteins in the supernatant. Glycosylation of 3D5 and expression of PglB were analysed by immunoblot following SDS-PAGE. Immunodetection was performed with anti-c-Myc monoclonal antibody (Calbiochem) and anti-glycan serum hR6 (S. Amber and M. Aebi, personal communication) to observe glycosylated 3D5. Immunodetection of *C. lari* PglB was performed with anti-HA antiserum (Santa Cruz). All experiments were performed at least in triplicate, and representative samples are shown.

**Mutagenesis study.** Mutant PglB was generated by the QuickChange method, and the resulting plasmids of all constructs were validated by DNA sequencing. The mutant PglB variants were cloned into pMLBAD as above and used in complementation assays.

**PglB purification.** The gene encoding PglB was cloned into a modified pBAD (Invitrogen) expression plasmid with a C-terminal decahistidine affinity tag fused to PglB, resulting in the plasmid pSF2. Owing to the applied cloning strategy, PglB carried the mutation K2E and the plasmid was confirmed by DNA sequencing (Microsynth).

PglB from *C. lari* was overexpressed from pSF2 in *E. coli* BL21-Gold (DE3) (Stratagene) cells in a 30 l fermentor (Infors). Cells were grown at 37 °C in Terrific Broth medium supplemented with 1% glycerol (w/v) to  $A_{600}$  of 10.0 before the culture was induced by the addition of 0.1% arabinose (w/v) for 2 h. All following steps were performed at 4 °C unless specified differently. Cells were harvested by centrifugation, re-suspended in 25 mM Tris-HCl, pH 8.0; 250 mM NaCl and disrupted in a M-110L microfluidizer (Microfluidics) at 15,000 p.s.i. chamber pressure. Membranes were pelleted by ultracentrifugation at 100,000g for 0.5 h. PglB was solubilized in 25 mM Tris-HCl, pH 8.0; 250 mM NaCl; 10% glycerol (v/v) and 1% *N*-dodecyl- $\beta$ -D-maltopyranoside (w/v) (DDM, Anatrace) for 1 h. All subsequent buffers contained DDM as detergent.

The supernatant was supplemented with 25 mM imidazole and loaded onto a NiNTA superflow affinity column (Qiagen), washed with 60 mM imidazole

before PglB was eluted with 200 mM imidazole. The protein was desalted into 10 mM MES-NaOH, pH 6.5; 100 mM NaCl; 0.5 mM EDTA; 3% glycerol (v/v); 3% polyethylene glycol 400 (v/v) and concentrated to 7–10 mg ml<sup>-1</sup> in an Amicon Ultra-15 concentrator (Millipore) with a molecular mass cutoff of 100 kDa.

**Native crystals.** The peptide Ac-DQNATF{4-NO<sub>2</sub>}-NH<sub>2</sub> was added to concentrated PglB to a final concentration of 0.75 mM, incubated for 0.5 h, and crystallized by vapour diffusion in sitting drops at 20 °C against a reservoir of 100 mM glycine, pH 9.4; 50 mM magnesium acetate; 6% dimethyl sulphoxide (DMSO) (v/v) and 23–34% (v/v) polyethylene glycol 400. The protein to reservoir volume ratio in the sitting drop was 2:1. Crystals typically appeared after 3–4 weeks and matured to full size within 6 weeks. Crystals were directly flash frozen by immersion in liquid nitrogen before data collection.

**Heavy-metal derivatives.** Native crystals were soaked for 30–60 min in 1 mM ethyl mercury phosphate (EMP) before back-soaking and flash-freezing by immersion in liquid nitrogen.

**Data collection.** Crystals belonged to the space group P2<sub>1</sub>2<sub>1</sub>2<sub>1</sub>, with one PglB-peptide complex in the asymmetric unit. Native data were collected at the micro-diffractometer beamline X06SA at the Swiss Light Source (SLS, Villigen) because not all sections of the crystals diffracted equally well. EMP2 and EMP3 derivative data sets (Supplementary Table 1) were collected at the same station, whereas EMP1 was collected at the high-resolution station of the same beam line. Data were processed and merged with XDS<sup>49</sup> or HKL2000 (HKL Research, Inc.) and anisotropic scaling/ellipsoid truncation was performed<sup>50</sup>.

**Structure determination.** The structure was determined using a combination of molecular replacement using the periplasmic domain of *C. jejuni* PglB (Protein Data Bank code 3AAG) as a search model and Phaser<sup>51</sup> on the one hand, and multiple isomorphous replacement with anomalous scattering (MIRAS) using SHARP (Global Phasing Limited) on the other. The process of phase calculation and model building (using O<sup>52</sup>) and refinement (using Phenix<sup>53</sup>) was iterated, starting with the periplasmic domain and extending into the best-ordered regions of the transmembrane domain (TM1–4 and TM10–13), followed by TM5–9. The locations of three cysteines in the transmembrane domain (indicated by Hg anomalous peaks) served as starting points for tracing, until very good density allowed placement of bulky residues, confirming the sequence register. The final structure excludes two disordered loops of PglB (residues 283–306 and residues 605–607) as well as the C-terminal polyhistidine tag. Data collection and refinement statistics are given in Supplementary Table 1.

48. Lefebvre, M. D. & Valvano, M. A. Construction and evaluation of plasmid vectors optimized for constitutive and regulated gene expression in *Burkholderia cepacia* complex isolates. *Appl. Environ. Microbiol.* **68**, 5956–5964 (2002).
49. Kabsch, W. XDS. *Acta Crystallogr. D* **66**, 125–132 (2010).
50. Strong, M. et al. Toward the structural genomics of complexes: Crystal structure of a PE/PPE protein complex from *Mycobacterium tuberculosis*. *Proc. Natl Acad. Sci. USA* **103**, 8060–8065 (2006).
51. McCoy, A. J. et al. Phaser crystallographic software. *J. Appl. Cryst.* **40**, 658–674 (2007).
52. Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. Improved methods for building protein models in electron-density maps and the location of errors in these models. *Acta Crystallogr. A* **47**, 110–119 (1991).
53. Adams, P. D. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D* **66**, 213–221 (2010).



# Black hole growth in the early Universe is self-regulated and largely hidden from view

Ezequiel Treister<sup>1,2</sup>, Kevin Schawinski<sup>3,4</sup>, Marta Volonteri<sup>5</sup>, Priyamvada Natarajan<sup>3,4,6,7</sup> & Eric Gawiser<sup>8</sup>

**The formation of the first massive objects in the infant Universe remains impossible to observe directly and yet it sets the stage for the subsequent evolution of galaxies<sup>1–3</sup>. Although some black holes with masses more than  $10^9$  times that of the Sun have been detected in luminous quasars less than one billion years after the Big Bang<sup>4,5</sup>, these individual extreme objects have limited utility in constraining the channels of formation of the earliest black holes; this is because the initial conditions of black hole seed properties are quickly erased during the growth process<sup>6</sup>. Here we report a measurement of the amount of black hole growth in galaxies at redshift  $z = 6–8$  (0.95–0.7 billion years after the Big Bang), based on optimally stacked, archival X-ray observations. Our results imply that black holes grow in tandem with their host galaxies throughout cosmic history, starting from the earliest times. We find that most copiously accreting black holes at these epochs are buried in significant amounts of gas and dust that absorb most radiation except for the highest-energy X-rays. This suggests that black holes grew significantly more during these early bursts than was previously thought, but because of the obscuration of their ultraviolet emission they did not contribute to the re-ionization of the Universe.**

The Chandra X-ray observatory is sensitive to photons in the energy range 0.5–8 keV, which in deep extragalactic observations probes predominantly accretion onto supermassive black holes<sup>7</sup>. Rapidly growing black holes are known to be surrounded by an obscuring medium, which can block most of the optical, ultraviolet and even soft X-ray photons<sup>8</sup>. With increasing redshift, at the earliest epochs, the photons observed by Chandra are emitted at intrinsically higher energies, and are therefore less affected by such absorption. Current X-ray observations have not been able to individually detect most of the first black hole growth events at  $z > 6$  (the first 950 million years after the Big Bang) thus far, except for the most luminous quasars<sup>9</sup> with X-ray luminosity  $L_X > 3 \times 10^{44} \text{ erg s}^{-1}$ . Whereas deep X-ray surveys do not cover enough volume at high redshift, current wide-area studies are simply not deep enough. Hence, the only way to obtain a detectable signal from more typical growing black holes is by adding the X-ray emission from a large number of sources at these redshifts; we pursue this strategy here.

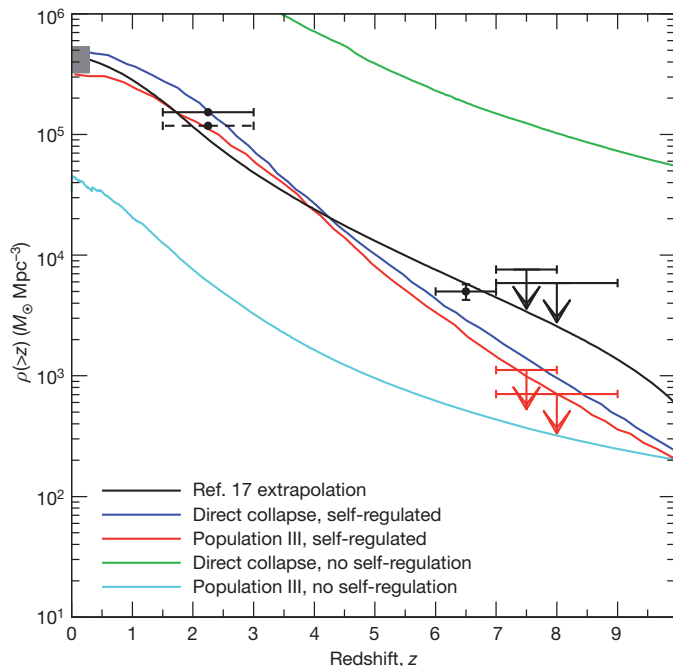
We start by studying the collective X-ray emission from the most distant galaxies known, at  $z \approx 6$  (ref. 10),  $z \approx 7$  (ref. 11) and  $z \approx 8$  (ref. 12), detected by the Wide Field Camera aboard the Hubble Space Telescope. These galaxies are as massive as today's galaxies (stellar mass<sup>13</sup>  $10^9–10^{11} M_\odot$ , where  $M_\odot$  is the solar mass), and they are thus likely to harbour substantial central black holes. None of the  $z > 6$  galaxies studied in this work are individually detected in the Chandra X-ray observations. However, we detect significant signals from a stack of 197 galaxies at  $z \approx 6$  in both the soft (0.5–2.0 keV; corresponding to 3.5–14 keV in the rest frame) and hard (2–8 keV; rest-frame 14–56 keV) X-ray bands independently. The detection in

the soft band is significant at the  $5\sigma$  level, and implies an average observed-frame luminosity of  $9.2 \times 10^{41} \text{ erg s}^{-1}$ , while in the hard band the stacked  $6.8\sigma$  signal corresponds to an average luminosity of  $8.4 \times 10^{42} \text{ erg s}^{-1}$ . For the sample of galaxies at  $z \approx 7$ , we obtain  $3\sigma$  upper limits for the average luminosity in the observed-frame soft and hard X-ray bands of  $4 \times 10^{42} \text{ erg s}^{-1}$  and  $2.9 \times 10^{43} \text{ erg s}^{-1}$ , respectively. Combining the  $z \approx 7$  and  $z \approx 8$  samples, the corresponding  $3\sigma$  upper limits are  $3.1 \times 10^{42} \text{ erg s}^{-1}$  and  $2.2 \times 10^{43} \text{ erg s}^{-1}$  in the observed-frame soft and hard X-ray bands, respectively.

A large difference, of a factor of  $\sim 9$ , is found between the stacked fluxes in the soft and hard X-ray bands at  $z \approx 6$ . This requires large amounts of obscuring material with high neutral hydrogen column densities ( $N_H > 1.6 \times 10^{24} \text{ cm}^{-2}$ ) to be present in a very high fraction of the accreting black holes in these galaxies, in order to explain the large deficit of soft X-ray photons. As this signal derives from the entire population, these results require almost all sources to be significantly obscured. This in turn implies that these growing black holes are obscured along most lines of sight, as is also observed in a small subset of nearby objects<sup>14</sup>. Such a high fraction of obscured sources at low luminosities is also observed at low redshifts<sup>15</sup>. This large amount of obscuration along all directions absorbs virtually all ultraviolet photons from growing black holes. Thus, regardless of the amount of accretion in these sources, these active galaxies cannot have contributed to the early re-ionization of the Universe. Alternatively, it cannot be claimed that rapid and efficient supermassive black hole growth in the high- $z$  Universe is implausible on the basis of any re-ionization constraints<sup>16</sup>. If most of the high-redshift black hole growth is indeed obscured, as suggested by our work, several current constraints on the lifetime and duty cycle of high- $z$  accreting black holes need to be revisited and revised.

Assuming that the X-ray emission is due to accretion onto the central black hole, the space density of mass accreted by black holes (in terms of solar masses per Mpc<sup>3</sup>) can be directly derived from the observed X-ray luminosity, as described in the Supplementary Information. Extrapolations of active galactic nuclei (AGN) luminosity functions<sup>17</sup> measured at significantly lower redshifts,  $z < 3$ , are consistent with the observed accreted black hole mass density at  $z > 6$ , as can be seen in Fig. 1. This directly leads to two further conclusions: the space density of low-luminosity ( $L_X < 10^{44} \text{ erg s}^{-1}$ ) sources does not evolve significantly from  $z \approx 1$  to  $z \approx 6–8$ , that is, over more than 5 billion years. Second, at higher luminosities, the extrapolation of lower-redshift AGN luminosity functions leads to an overestimate of the observed source density in optical surveys<sup>18</sup>. This discrepancy can be resolved if the shape of the AGN luminosity function evolves strongly, in the sense that there are relatively fewer high-luminosity AGN at  $z > 6$  in comparison to the  $z < 3$  population. Another possibility is that the number of obscured sources, relative to unobscured quasars, increases with redshift, such that most of the highly obscured systems are systematically missed in these optical studies. This is strongly supported by

<sup>1</sup>Institute for Astronomy, 2680 Woodlawn Drive, University of Hawaii, Honolulu, Hawaii 96822, USA. <sup>2</sup>Universidad de Concepción, Departamento de Astronomía, Casilla 160-C, Concepción, Chile. <sup>3</sup>Yale Center for Astronomy and Astrophysics, PO Box 208121, New Haven, Connecticut 06520, USA. <sup>4</sup>Department of Physics, Yale University, PO Box 208121, New Haven, Connecticut 06520, USA. <sup>5</sup>Department of Astronomy, University of Michigan, Ann Arbor, Michigan 48109, USA. <sup>6</sup>Department of Astronomy, Yale University, PO Box 208101, New Haven, Connecticut 06520, USA. <sup>7</sup>Institute for Theory and Computation, Harvard University, 60 Garden Street, Cambridge, Massachusetts 02138, USA. <sup>8</sup>Department of Physics and Astronomy, Rutgers University, 136 Frelinghuysen Road, Piscataway, New Jersey 08854, USA.



**Figure 1 | Accreted black hole mass density as a function of redshift.** Grey rectangle (top left) shows the range of values allowed by observations of  $z \approx 0$  galaxies<sup>28</sup>. The data points at  $z \approx 2$  correspond to the values obtained from Chandra observations of X-ray detected AGN (dashed error bars) and luminous infrared galaxies<sup>29</sup> (solid error bars), while the measurement at  $z \approx 6-7$  and the upper limits (downward arrows) at  $z = 7-9$  show the results described in this work (red and black data points from the observed-frame soft and hard X-ray band observations, respectively). Vertical error bars, 1 s.d.; horizontal error bars, bin size. Black solid line, evolution of the accreted black hole mass density inferred from the extrapolation of AGN luminosity functions measured at lower redshifts<sup>17</sup>. We overplot the predictions of black hole and galaxy evolution models<sup>30</sup> for non-regulated growth of population-III star remnants (cyan line) and direct-collapse seeds (green line). The red and blue lines show the predicted black hole mass density if self-regulation is incorporated.

observations of quasars at lower redshifts,  $z < 3$  (ref. 19). We cannot rule out either of these scenarios at present owing to the relatively small cosmological volume studied, in which the extremely rare high-luminosity AGN are absent.

Our measurements and upper limits for the accreted black hole mass density up to  $z \approx 8.5$  ( $\sim 600$  million years after the Big Bang) constrain the nature of black hole growth in the early Universe. Two critical issues for AGN and the supermassive black holes powering them are how the first black holes formed, and how they subsequently grew by accreting mass while shining as AGN. The strong local correlation between black hole mass and galaxy bulge mass observed at  $z \approx 0$  (refs 20, 21) is interpreted as evidence for self-regulated black hole growth and galaxy–black hole co-evolution<sup>1,22</sup>. This is currently the default assumption for most galaxy formation and evolution models<sup>23,24</sup>.

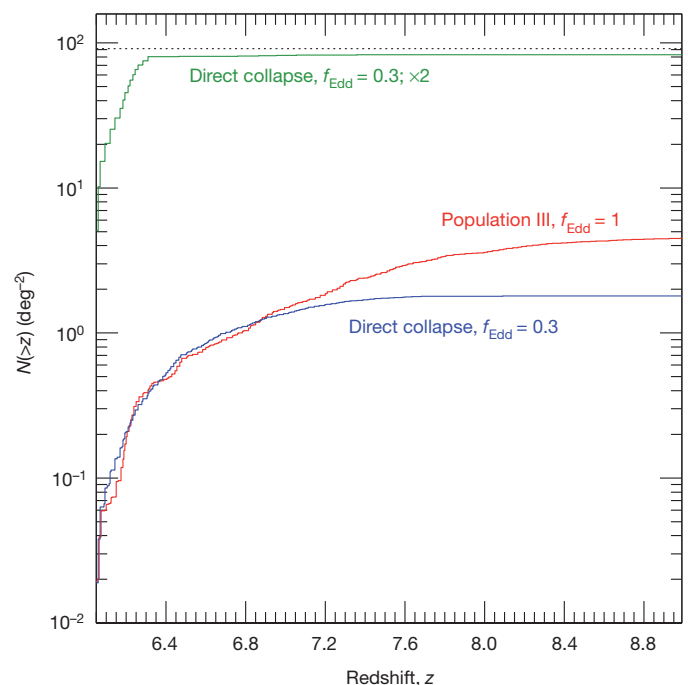
The origin of the initial ‘seed’ black holes remains an unsolved problem at present. Two channels to form these seeds have been proposed: compact remnants of the first stars, the so-called population III stars<sup>25</sup>, which generate seeds with masses  $\sim 10-10^3 M_\odot$ ; and the direct gravitational collapse of gas-rich pre-galactic disks, which leads to significantly more massive seeds with masses in the range  $10^3-10^5 M_\odot$  (refs 26, 27). By modelling, we find that the masses of seeds that form from direct collapse are correlated to properties of the dark matter halo and hence properties of the galaxy that will assemble subsequently.

To interpret our finding, we explore a theoretical framework for the cosmic evolution of supermassive black holes. We follow the formation and evolution of black holes through dedicated Monte Carlo merger tree simulations. Each model is constructed by tracing the merger hierarchy of dark matter haloes in the mass range  $10^{11}-10^{15}$

$M_\odot$  backwards to  $z = 20$ , using an extended Press and Schechter algorithm<sup>3</sup>. The haloes are then seeded with black holes and their evolution is tracked forward to the present time. Following a major merger (defined as a merger between two haloes with mass ratio  $> 0.1$ ), supermassive black holes accrete efficiently an amount of mass that is set by a ‘self-regulated’ model (where the accreted mass scales with the fourth power of the host halo circular velocity and is normalized to reproduce the observed local correlation between supermassive black hole mass and velocity dispersion) or by an ‘unregulated’ model, where the supermassive black hole simply doubles in mass at each accretion episode. See Supplementary Information for additional details.

Our observational results provide strong support for the existence of a correlation between supermassive black holes and their hosts out to the highest redshifts. In Fig. 1, we compare both unregulated and self-regulated black hole growth models with our observations, and find that physically motivated self-regulation growth models are highly favoured at all redshifts, even in the very early Universe. Unregulated models (for instance, those in which black holes just double in mass at each major merger) are strongly disfavoured by the data. This indicates that even in the first episodes of black hole growth, there is a fundamental link between galaxy and black hole mass assembly.

As shown in Fig. 1, once a standard prescription for self-regulation (as described above) is incorporated, both seed models are consistent with our current high- $z$  observations. Detection of an unbiased population of sources at these early epochs is the one metric that we have in the foreseeable future to distinguish between these two scenarios for the origin of supermassive black holes in the Universe. In Fig. 2, we present the predicted cumulative source counts at  $z > 6$  for the models studied here. On the basis of these models, ultra-deep X-ray and near-infrared surveys covering at least  $\sim 1$  degree<sup>2</sup> are required to constrain



**Figure 2 | Cumulative number of sources as a function of redshift for individual X-ray detections.** This calculation assumes the X-ray flux limit of the 4 Ms CDF-S Chandra observations. The horizontal dotted line shows the number density required to individually detect one source in the area considered in this work at  $z > 7$ . Models are described in the Supplementary Information and labelled in the figure. Red, population III, Eddington fraction ( $f_{\text{Edd}} = 1$ ); blue, direct collapse,  $f_{\text{Edd}} = 0.3$ ; black, direct collapse,  $f_{\text{Edd}} = 0.3, \times 2$ . We note that model population III,  $f_{\text{Edd}} = 1, \times 2$  has no detectable source. To distinguish between these models for early black hole formation will require a deep multiwavelength survey covering at least  $\sim 1$  degree<sup>2</sup>.

the formation of the first black hole seeds. This will probably require the use of the next generation of space-based observatories, such as the James Webb Space Telescope and the International X-ray Observatory.

Received 3 December 2010; accepted 7 April 2011.

1. Silk, J. & Rees, M. J. Quasars and galaxy formation. *Astron. Astrophys.* **331**, L1–L4 (1998).
2. Barkana, R. & Loeb, A. In the beginning: the first sources of light and the reionization of the universe. *Phys. Rep.* **349**, 125–238 (2001).
3. Volonteri, M., Haardt, F. & Madau, P. The assembly and merging history of supermassive black holes in hierarchical models of galaxy formation. *Astrophys. J.* **582**, 559–573 (2003).
4. Fan, X. *et al.* High-redshift quasars found in Sloan Digital Sky Survey commissioning data. IV. Luminosity function from the full equatorial stripe sample. *Astron. J.* **121**, 54–65 (2001).
5. Willott, C. J. *et al.* Four quasars above redshift 6 discovered by the Canada-France High-*z* Quasar Survey. *Astron. J.* **134**, 2435–2450 (2007).
6. Volonteri, M. & Rees, M. J. Quasars at  $z=6$ : the survival of the fittest. *Astrophys. J.* **650**, 669–678 (2006).
7. Brandt, W. N. & Hasinger, G. Deep extragalactic X-ray surveys. *Annu. Rev. Astron. Astrophys.* **43**, 827–859 (2005).
8. Lawrence, A. & Elvis, M. Obscuration and the various kinds of Seyfert galaxies. *Astrophys. J.* **256**, 410–426 (1982).
9. Shemmer, O. *et al.* Chandra observations of the highest redshift quasars from the Sloan Digital Sky Survey. *Astrophys. J.* **644**, 86–99 (2006).
10. Bouwens, R. J., Illingworth, G. D., Blakeslee, J. P. & Franx, M. Galaxies at  $z\sim 6$ : the UV luminosity function and luminosity density from 506 HUDF, HUDF Parallel ACS Field, and GOODS i-dropouts. *Astrophys. J.* **653**, 53–85 (2006).
11. Bouwens, R. J. *et al.* UV luminosity functions from 113  $z\sim 7$  and  $z\sim 8$  Lyman-break galaxies in the ultra-deep HUDF09 and wide-area ERS WFC3/IR observations. Preprint at (<http://arXiv.org/abs/1006.4360>) (2010).
12. Bouwens, R. J. *et al.* Discovery of  $z\sim 8$  galaxies in the Hubble Ultra Deep Field from ultra-deep WFC3/IR observations. *Astrophys. J.* **709**, L133–L137 (2010).
13. Labbé, I., Bouwens, R., Illingworth, G. D. & Franx, M. Spitzer IRAC confirmation of  $z_{850}$ -dropout galaxies in the Hubble Ultra Deep Field: stellar masses and ages at  $z\sim 7$ . *Astrophys. J.* **649**, L67–L70 (2006).
14. Ueda, Y. *et al.* Suzaku observations of active galactic nuclei detected in the Swift BAT survey: discovery of a “new type” of buried supermassive black holes. *Astrophys. J.* **664**, L79–L82 (2007).
15. Sazonov, S., Revnivtsev, M., Krivonos, R., Churazov, E. & Sunyaev, R. Hard X-ray luminosity function and absorption distribution of nearby AGN: INTEGRAL all-sky survey. *Astron. Astrophys.* **462**, 57–66 (2007).
16. Loeb, A. The race between stars and quasars in reionizing cosmic hydrogen. *J. Cosmol. Astropart. Phys.* **3**, 022 (2009).
17. Treister, E., Urry, C. M. & Virani, S. The space density of Compton thick AGN and the X-ray background. *Astrophys. J.* **696**, 110–120 (2009).
18. Willott, C. J. *et al.* The Canada-France High-*z* Quasar Survey: nine new quasars and the luminosity function at redshift 6. *Astron. J.* **139**, 906–918 (2010).
19. Treister, E. *et al.* Major galaxy mergers and the growth of supermassive black holes in quasars. *Science* **328**, 600–602 (2010).
20. Ferrarese, L. & Merritt, D. A fundamental relation between supermassive black holes and their host galaxies. *Astrophys. J.* **539**, L9–L12 (2000).
21. Gebhardt, K. *et al.* A relationship between nuclear black hole mass and galaxy velocity dispersion. *Astrophys. J.* **539**, L13–L16 (2000).
22. King, A. Black holes, galaxy formation, and the  $M_{\text{BH}}-\sigma$  relation. *Astrophys. J.* **596**, L27–L29 (2003).
23. Wyithe, J. S. B. & Loeb, A. Self-regulated growth of supermassive black holes in galaxies as the origin of the optical and X-ray luminosity functions of quasars. *Astrophys. J.* **595**, 614–623 (2003).
24. Hopkins, P. F. *et al.* A unified, merger-driven model of the origin of starbursts, quasars, the cosmic X-ray background, supermassive black holes, and galaxy spheroids. *Astrophys. J.* **163**, 1–49 (2006).
25. Madau, P. & Rees, M. J. Massive black holes as population III remnants. *Astrophys. J.* **551**, L27–L30 (2001).
26. Bromm, V. & Loeb, A. Formation of the first supermassive black holes. *Astrophys. J.* **596**, 34–46 (2003).
27. Lodato, G. & Natarajan, P. Supermassive black hole formation during the assembly of pregalactic discs. *Mon. Not. R. Astron. Soc.* **371**, 1813–1823 (2006).
28. Shankar, F., Weinberg, D. H. & Miralda-Escudé, J. Self-consistent models of the AGN and black hole populations: duty cycles, accretion rates, and the mean radiative efficiency. *Astrophys. J.* **690**, 20–41 (2009).
29. Treister, E., Urry, C. M., Schawinski, K., Cardamone, C. N. & Sanders, D. B. Heavily obscured active galactic nuclei in high-redshift luminous infrared galaxies. *Astrophys. J.* **722**, L238–L243 (2010).
30. Volonteri, M. Formation of supermassive black holes. *Astron. Astrophys. Rev.* **18**, 279–315 (2010).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank T. Goto, M. Urry and D. Sanders for conversations. Support for the work of E.T. and K.S. was provided by NASA through Chandra/Einstein Post-doctoral Fellowship Awards. M.V. acknowledges support from the Smithsonian Astrophysical Observatory. P.N. acknowledges support via a Guggenheim Fellowship from the John Simon Guggenheim Foundation. The work of E.G. was partially funded by the NSF.

**Author Contributions** E.T. started the project, collected the galaxy samples, performed the X-ray stacking calculations and wrote the majority of the text. K.S. helped to collect the galaxy sample studied here, and contributed to the conception of the project and the analysis and interpretation of the results. M.V. and P.N. created the black hole growth models, computed the contribution of these sources to the re-ionization of the Universe and contributed extensively to the theoretical interpretation of the observational results. E.G. developed the optimal X-ray stacking formalism and worked with E.T. to implement it on these data. All authors discussed the results and contributed to the writing of the manuscript.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at [www.nature.com/nature](http://www.nature.com/nature). Correspondence and requests for materials should be addressed to E.T. ([treister@ifa.hawaii.edu](mailto:treister@ifa.hawaii.edu)).



# Zero outward flow velocity for plasma in a heliosheath transition layer

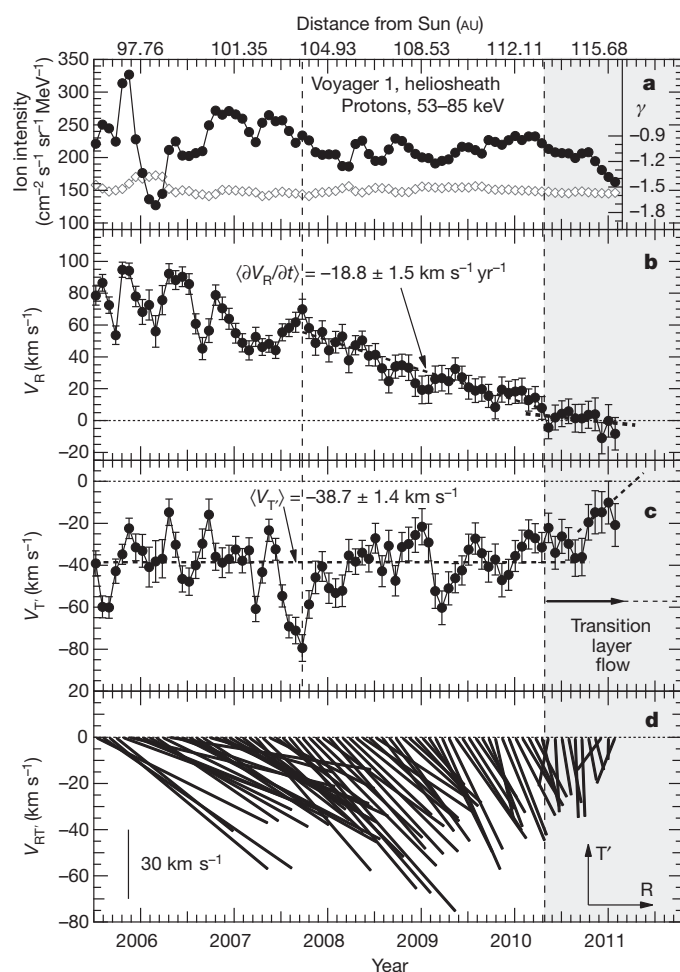
Stamatios M. Krimigis<sup>1,2</sup>, Edmond C. Roelof<sup>1</sup>, Robert B. Decker<sup>1</sup> & Matthew E. Hill<sup>1</sup>

Voyager 1 has been in the reservoir of energetic ions and electrons that constitutes the heliosheath since it crossed the solar wind termination shock<sup>1–3</sup> on 16 December 2004 at a distance from the Sun of 94 astronomical units (1 AU =  $1.5 \times 10^8$  km). It is now  $\sim 22$  AU past the termination shock crossing<sup>4</sup>. The bulk velocity of the plasma in the radial-transverse plane has been determined<sup>5</sup> using measurements of the anisotropy of the convected energetic ion distribution<sup>6</sup>. Here we report that the radial component of the velocity has been decreasing almost linearly over the past three years, from  $\sim 70$  km s<sup>–1</sup> to  $\sim 0$  km s<sup>–1</sup>, where it has remained for the past eight months. It now seems that Voyager 1 has entered a finite transition layer of zero-radial-velocity plasma flow, indicating that the spacecraft may be close to the heliopause, the border between the heliosheath and the interstellar plasma. The existence of a flow transition layer in the heliosheath contradicts current

predictions<sup>7</sup>—generally assumed by conceptual models—of a sharp discontinuity at the heliopause.

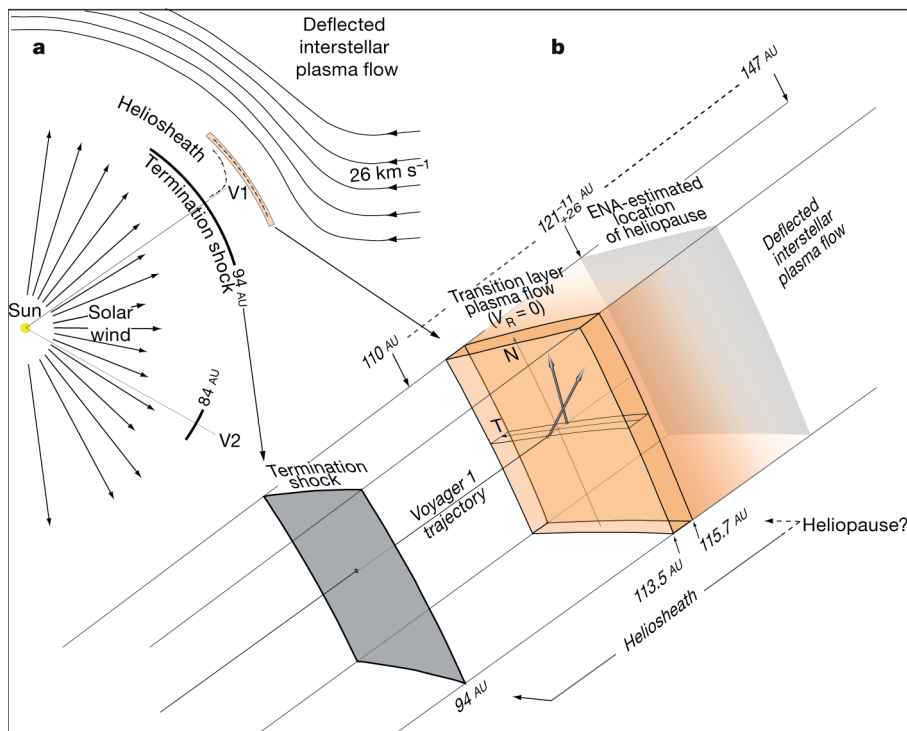
Figure 1a shows that there has been little change in the intensity of lower-energy ions at the position of Voyager 1 since April 2007. Indeed, a flat intensity profile and near-constant power-law energy spectrum ( $j \propto E^{-\gamma}$ ) have been persistently observed at all ion energies from  $\sim 40$  keV to 2 MeV, as is indicated by the constancy of the exponent  $\gamma$  (Fig. 1a, diamonds), which has been computed at 53–85 keV but is typical of the entire energy range. We note for later reference the small but steep intensity decrease during the last three months of 2010 observed for all low-energy ions (because of the nearly constant spectral shape).

The heliosheath plasma flow velocity as estimated from ion intensity anisotropies measured using LECP<sup>8</sup> is shown in Fig. 1b–d. In Fig. 1b, the radial component,  $V_R$ , undergoes a long, steady, nearly linear decline beginning at  $\sim 70$  km s<sup>–1</sup> and reaching  $\sim 0$  km s<sup>–1</sup> in April 2010 with an abrupt change in slope to nearly zero. This remarkable near-zero radial velocity continued until at least February 2011 (the spacecraft velocity of 17 km s<sup>–1</sup> has been subtracted). We regard the abrupt change in  $\langle \partial V_R / \partial t \rangle$  in April 2010 (from negative to zero) as an indication that Voyager 1 had entered a finite ‘transition layer’, terminology suggested in ref. 9. The more conventional conception of the heliopause has been that of a surface (separatrix) between the heated solar wind plasma of the heliosheath and the cold interstellar plasma of the very local near-interstellar medium, such that the theoretically



**Figure 1 | Directional velocity measurements from Voyager 1 at the edge of the heliosphere.** The Low Energy Charged Particle (LECP) instrument on Voyager 1 provides angular information via a mechanically stepped platform in eight 45° sectors. **a–c**, The velocity components of plasma flow (**b**, **c**) are calculated, using a method published elsewhere<sup>4,6</sup>, from the directional measurements of the 53–85-keV ion intensities, whose scan-average intensity and power-law spectral index ( $\gamma$ ) are shown in **a**. Briefly, spacecraft-frame intensities,  $j(\phi)$ , are represented by a second-order Fourier series in the scan angle,  $\phi$  ( $0 < \phi < 2\pi$ ), where  $j(\phi) = A_0 + A_1 \sin(\phi - \phi_1) + A_2 \sin(\phi - \phi_2)$ . The parameters  $A_0$ ,  $(A_1, \phi_1)$  and  $(A_2, \phi_2)$ , which generate the harmonic anisotropy amplitudes,  $\xi_1$  and  $\xi_2$ , are determined by a least-squares fit<sup>6</sup> to intensities in sectors 1–7. For  $\xi_2 \ll \xi_1$ ,  $\xi_1 \approx 2(\gamma + 1)V/v$ , implying that  $V = v\xi_1/2(\gamma + 1)$ , where  $v$  is the velocity of the energetic particles. The velocity is resolved into components in the Voyager 1 R–T’ instrument scan plane, which is rotated 20° anticlockwise about the radial (+R) direction from the R–T plane in the conventional R–T–N heliographic polar coordinates in which the transverse direction (+T) is that of planetary motion around the Sun. The error bars (in **a–c**; smaller than the point sizes in **a**) are computed from the Poisson standard deviations in the directional rates. Plots of the adjacent channels (40–53 keV and 80–139 keV) show nearly identical results<sup>4</sup>. Such agreement independent of particle energy is an assurance that the energetic ion distribution is indeed being advected with the plasma bulk velocity. A least-squares fit to the last ten points of  $V_R$  (**b**, heavy dashed line) gives an average velocity of  $\langle V_R \rangle = 1.0 \pm 2.4$  km s<sup>–1</sup> with an average slope of  $\langle \partial V_R / \partial t \rangle = -6.1 \pm 12.1$  km s<sup>–1</sup> yr<sup>–1</sup>, such that both mean and slope are statistically consistent with zero. In **c**, a least-squares fit to the transverse component,  $V_{T'}$ , is shown by the dashed line. **d**, The plasma flow velocity in the R–T’ plane represented as vectors, the head giving the velocity ( $V_{RT'}$ ) and the tail being located at the observation time along the time axis.

<sup>1</sup>Applied Physics Laboratory, The Johns Hopkins University, Laurel, Maryland 20723, USA. <sup>2</sup>Academy of Athens, Athens 11527, Greece.



**Figure 2 | The heliosphere and its boundaries in the general direction of Voyager 1.** **a**, Large-scale illustration depicting the solar wind radial flow inside the termination shock and the expected deflection of the heliosheath flow between the termination shock and the heliopause. The heliopause marks the (theoretical) boundary between the heliosheath plasma flow and the much denser, colder and slower flow of interstellar plasma being deflected around the heliosheath. V1, Voyager 1; V2, Voyager 2. **b**, Scale drawing of the unexpected transition layer that Voyager 1 has encountered within the heliosheath. For illustrative purposes, we have assigned a value ( $V_N \approx 40 \text{ km s}^{-1}$ ) to the (unmeasured) meridional velocity component throughout the layer. Voyager 1 measurements have also revealed the thickness of the transition layer, a possible location for the heliopause and the consistency of the range of locations of the

heliopause estimated from energetic-neutral-atom (ENA) images made by NASA's Cassini spacecraft. For the last, we used the simplified formula<sup>10</sup>  $j_{\text{ENA}} = \sigma |dr n_{\text{ion}} \approx \sigma L n_{\text{ion}}$ , which relates the ENA intensity,  $j_{\text{ENA}}$ , at Cassini to the ion intensity,  $j_{\text{ion}}$ , along the line of sight (distance,  $L$ ) between the termination shock and the heliopause through the image pixel containing Voyager 1. Here  $n = 0.1 \text{ cm}^{-3}$  is the cold interstellar neutral density and  $\sigma$  is the (energy-dependent) charge-exchange cross-section for proton–hydrogen collisions. We showed<sup>11</sup> that the ENA and ion spectra could be brought into agreement at Voyager 2 with a heliosheath thickness of  $L_2 = 54_{-30}^{+15} \text{ AU}$ , whereas the same normalization procedure applied to Voyager 1 results in  $L_1 = 27_{-26}^{+11} \text{ AU}$ .

expected gradual reduction of  $V_R$  (due to the rotation of the flow so as to parallel the heliopause) would end with  $V_R$  asymptotically (not abruptly) going to zero, and then only at the heliopause itself.

Figure 1c shows that the transverse velocity,  $V_T$ , fluctuated around a mean value of  $\langle V_T \rangle \approx -40 \text{ km s}^{-1}$  for  $\sim 5$  yr until near the end of 2010, when a trend towards zero began (Fig. 1c, last four points). The swing of the velocity vector (Fig. 1d) from radial to transverse was completed by June 2010. Because  $V_R$  has remained near zero and  $V_T$  seems to be tending to zero at the end of 2010, any remaining flow will have to be in the unobserved meridional component,  $V_N$ . For an axisymmetric heliopause, it is expected that  $V_N \approx 30 \text{ km s}^{-1}$  for the meridional flow of the deflected distant interstellar plasma ( $V = 26 \text{ km s}^{-1}$ ). The observed flow pattern therefore could be consistent with Voyager 1 now being in the deflected interstellar plasma flow; that is, Voyager 1 may have crossed the heliopause, having passed through the transition layer ( $V_R = 0$ ).

We have interpreted the trends in  $V_R$  after July 2007 as a spatial phenomenon, as it is unlikely to be all or in part temporal. We cannot construct a reasonable scenario dominated by temporal variations for the monotonic time dependence of  $V_R$ , for example by ascribing it to a continually accelerating inward motion of the termination shock for almost 3 yr that brings the heliosheath at 115 AU to a dead stop for at least eight months throughout a region  $\sim 2.5 \text{ AU}$  thick (the distance travelled by Voyager 1 in eight months). (All absolute distances are measured relative to the Sun.) The effects of any inward motion of the termination shock (such as to reduce  $V_R$ ) would take at least 1 yr to propagate the  $\sim 20 \text{ AU}$  to Voyager 1, and it is hard to imagine how it

could produce the sharp change in  $\langle \partial V_R / \partial t \rangle$  to zero observed in April 2010.

The spatial relationships in the transition layer of the plasma flow measurements along the radial trajectory of Voyager 1 (heliographic latitude,  $36^\circ \text{ N}$ ) are depicted in Fig. 2. If  $V_R$  remains zero, and the outer end of the flow transition layer is really the heliopause, then the Voyager 1 observations demand that the orientation of the heliopause is normal to the radial trajectory of Voyager 1 (regardless of the values of  $V_T$  and  $V_N$ ). The measured transition region ( $V_R \approx 0$ ) extends from 113.5 AU to at least 115.7 AU. To relate this location to estimates of the thickness,  $L$ , of the heliosheath, we have estimated  $L$  from the ENA all-sky images from the Cassini Ion and Neutrals Camera<sup>10,11</sup>, whose energy range overlaps that of LEP. Using the ENA data for Voyager 1 and the same method of analysis<sup>11</sup>, we compute a heliosheath thickness of  $L_1 = 27_{-26}^{+11} \text{ AU}$ . Assuming that the termination shock is still where Voyager 1 crossed it, at 94 AU, the estimated radius of the heliopause along the trajectory of Voyager 1 should be 121 AU, which is not inconsistent with our suggestion that Voyager 1 is actually crossing the heliopause if  $V_R$  and  $V_T$  remain near zero beyond 116 AU. This is why we called attention to the small ( $\sim 27\%$ ) decrease in the 0.04–2-MeV ions (mentioned above, in the discussion of Fig. 1a) that commenced just when  $V_T$  began increasing to zero. If the ion intensity decrease continues, we would interpret it as a draining away of the heliosheath's energetic ion population into the downstream interstellar plasma flow. However, recent activity<sup>12</sup> at Voyager 2 suggests that it may be a global heliospheric response to a changing magnetic configuration at the Sun.

We must remember that Voyager 1, because it has no operational instruments that can measure them, is 'blind' to the particles that produce the suprathermal ( $\sim 1$ –20-keV) pressure in the heliosheath that dominates the dynamics within the heliosheath and hence also the cross-heliopause force balance with the stress applied by the interstellar magnetic field<sup>13</sup>. It is this 'unseen' population that must be producing the structure we measure as the transition layer, so there remains the possibility that the heliopause may be completely different from anything that has been suggested by contemporary theory<sup>7,14,15</sup>. It would not be the first time that the Voyager observations have surprised us.

Received 21 January; accepted 11 April 2011.

1. Decker, R. B. *et al.* Voyager 1 in the foreshock, termination shock, and heliosheath. *Science* **309**, 2020–2024 (2005).
2. Burlaga, L. F. *et al.* Crossing the termination shock into the heliosheath: magnetic fields. *Science* **309**, 2027–2029 (2005).
3. Stone, E. C. *et al.* Voyager 1 explores the termination shock region and the heliosheath beyond. *Science* **309**, 2012–2020 (2005).
4. Decker, R. B., Krimigis, S. M., Roelof, E. C. & Hill, M. E. Variations of low-energy ion distributions measured in the heliosheath. *AIP Conf. Proc.* **1302**, 51–57 (2010).
5. Krimigis, S. M. *et al.* The Low Energy Charged Particle (LECP) experiment on the Voyager spacecraft. *Space Sci. Rev.* **21**, 329–354 (1977).
6. Krimigis, S. M. *et al.* Voyager 1 exited the solar wind at a distance of 85 au from the Sun. *Nature* **426**, 45–48 (2003).
7. Florinski, V. *et al.* The dynamic heliosphere: outstanding issues. *Space Sci. Rev.* **143**, 57–83 (2009).
8. Scientific exploration: What a long, strange trip it's been. *Nature* **454**, 24–25 (2008).
9. Suess, S. T. The heliopause. *Rev. Geophys.* **28**, 97–115 (1990).
10. Krimigis, S. M., Mitchell, D. G., Roelof, E. C., Hsieh, K. C. & McComas, D. J. Imaging the interaction of the heliosphere with the interstellar medium from Saturn with Cassini. *Science* **326**, 971–973 (2009).
11. Krimigis, S. M., Mitchell, D. G., Roelof, E. C. & Decker, R. B. ENA ( $E > 5$  keV) images from Cassini and Voyager "ground truth": suprathermal particle pressure in the heliosheath. *AIP Conf. Proc.* **1302**, 79–85 (2010).
12. Decker, R. B., Roelof, E. C., Krimigis, S. M. & Hill, M. E. in *Physics of the Heliosphere: A 10-Year Retrospective* (eds Heerikhuisen, J., Li, G. & Zank, G.) (10th Annu. Internat. Astrophys. Conf., American Institute of Physics, 2011).
13. Roelof, E. C. *et al.* Implications of generalized Rankine-Hugoniot conditions for the PUI population at the Voyager 2 termination shock. *AIP Conf. Proc.* **1302**, 133–141 (2010).
14. Izmodenov, V. V. *et al.* Kinetic-gasdynamic modeling of the heliospheric interface. *Space Sci. Rev.* **146**, 329–351 (2009).
15. Zank, G. P. Physics of the solar wind-local interstellar medium interaction: role of magnetic fields. *Space Sci. Rev.* **146**, 295–327 (2009).

**Acknowledgements** This work was supported at The Johns Hopkins University Applied Physics Laboratory by NASA. We are grateful to J. Aiello (for assistance with our graphical presentation) and R. McNutt (for a historical summary of heliosheath terminology).

**Author Contributions** S.M.K. contributed most of the text; E.C.R. contributed to the text and provided theory interpretation; and R.B.D. performed the data analysis with the assistance of M.E.H.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at [www.nature.com/nature](http://www.nature.com/nature). Correspondence and requests for materials should be addressed to S.M.K. (tom.krimigis@jhuapl.edu).



# Emerging local Kondo screening and spatial coherence in the heavy-fermion metal $\text{YbRh}_2\text{Si}_2$

S. Ernst<sup>1</sup>, S. Kirchner<sup>1,2</sup>, C. Krellner<sup>1</sup>, C. Geibel<sup>1</sup>, G. Zwicknagl<sup>3</sup>, F. Steglich<sup>1</sup> & S. Wirth<sup>1</sup>

The entanglement of quantum states is both a central concept in fundamental physics and a potential tool for realizing advanced materials and applications. The quantum superpositions underlying entanglement are at the heart of the intricate interplay of localized spin states and itinerant electronic states that gives rise to the Kondo effect in certain dilute magnetic alloys<sup>1</sup>. In systems where the density of localized spin states is sufficiently high, they can no longer be treated as non-interacting; if they form a dense periodic array, a Kondo lattice may be established<sup>1</sup>. Such a Kondo lattice gives rise to the emergence of charge carriers with enhanced effective masses, but the precise nature of the coherent Kondo state responsible for the generation of these heavy fermions remains highly debated<sup>1–3</sup>. Here we use atomic-resolution tunnelling spectroscopy to investigate the low-energy excitations of a generic Kondo lattice system,  $\text{YbRh}_2\text{Si}_2$ . We find that the hybridization of the conduction electrons with the localized  $4f$  electrons results in a decrease in the tunnelling conductance at the Fermi energy. In addition, we observe unambiguously the crystal-field excitations of the  $\text{Yb}^{3+}$  ions. A strongly temperature-dependent peak in the tunnelling conductance is attributed to the Fano resonance<sup>4,5</sup> resulting from tunnelling into the coherent heavy-fermion states that emerge at low temperature. Taken together, these features reveal how quantum coherence develops in heavy  $4f$ -electron Kondo lattices. Our results demonstrate the efficiency of real-space electronic structure imaging for the investigation of strong electronic correlations<sup>6,7</sup>, specifically with respect to coherence phenomena, phase coexistence and quantum criticality.

The heavy charge carriers of Kondo lattice systems form from a lattice of magnetic ions coupled to itinerant electrons. The strongly correlated quantum states in such heavy-fermion metals are characterized by enhanced effective charge-carrier masses,  $m^*$ , and can give rise to a wide spectrum of emergent behaviour. These include various types of normal-metallic paramagnetic, magnetically ordered and superconducting phases, the latter two being either competing or coexisting. The resulting energy scales are typically small enough to allow these systems to be ‘tuned’ through different phases. Consequently, heavy-fermion metals are so far the only electronic systems in which the existence of a continuous phase transition at zero temperature, a so-called quantum critical point, has been established<sup>8</sup>.

However, the precise nature of the coherent Kondo state in Kondo lattice systems and the involved energy scales that are accessed on lowering the temperature,  $T$ , remain enigmatic<sup>2,3</sup>. This is mainly due to the lack of reliable local probes at these low energy scales. For single (or a few) magnetic adatoms on noble-metal surfaces, the formation of a Kondo state has been addressed by scanning tunnelling microscopy<sup>9–13</sup> (STM). The hybridization of itinerant and localized degrees of freedom is reflected in the observation of a so-called Fano resonance of the tunnelling conductance<sup>4,5</sup>. But to address the issue of evolving spatial coherence<sup>1</sup>, Kondo lattice systems need to be investigated. Recently, STM work on  $\text{URu}_2\text{Si}_2$  again revealed a Fano resonance and, additionally, a gap-like structure in the ‘hidden-order’ phase

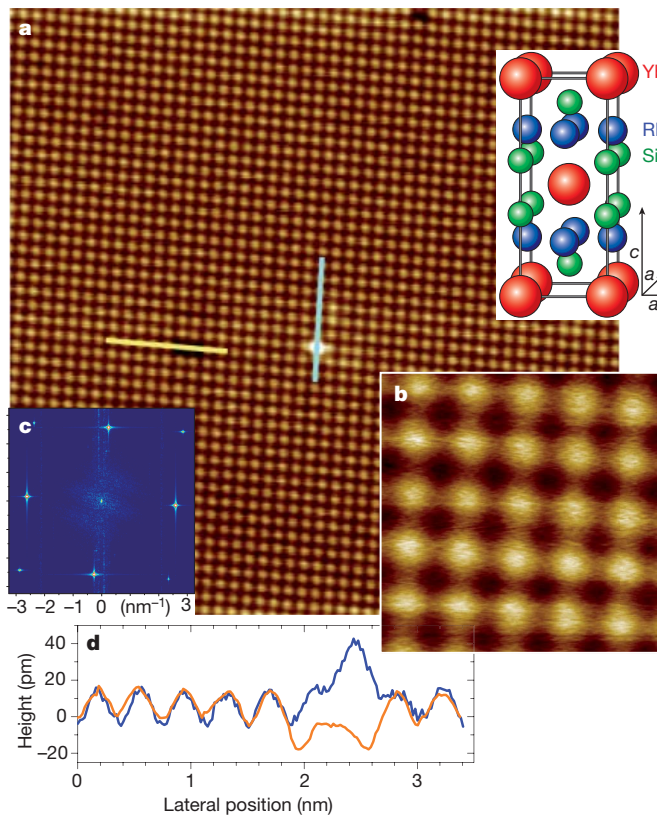
below a certain critical temperature,  $T_0 = 17.5$  K (refs 6, 7). In ref. 6 this gap is related, by means of imaging heavy-quasiparticle interference, to a rapid splitting of a light band into two new heavy-fermion bands, whereas in ref. 7 it is ascribed to a mean-field-type order parameter. However, the concept of Kondo screening in actinide-based compounds is debated<sup>14</sup> mainly because of the more delocalized character of  $5f$  electrons relative to  $4f$  electrons<sup>15</sup>.

We have therefore investigated a prototype Kondo lattice system,  $\text{YbRh}_2\text{Si}_2$ . The local character of the  $4f$  electrons is demonstrated, for example, by the observation of a well-defined crystal-electric-field (CEF) splitting of the  $J = 7/2$  Hund’s rule multiplet<sup>16</sup>, where  $J$  denotes total angular momentum. Moreover, the Kondo screening involving the  $4f$  moments and the conduction electrons has been well established<sup>17</sup> by the observation of a strong increase in  $m^*$  (up to the order of  $10^3$  free-electron masses). For  $\text{YbRh}_2\text{Si}_2$ , the single-ion Kondo temperature of the thermally excited CEF states<sup>18</sup> is, according to transport measurements,  $T_K = 80$ – $100$  K (refs 19, 20); and entropy<sup>19</sup> and thermopower<sup>20</sup> results reveal a single-ion Kondo temperature of  $\sim 30$  K for the CEF ground state<sup>18</sup>. Exploring the formation of Kondo lattice coherence at a temperature  $T_L$ , which is expected to occur below  $T_K$ , is one of the main aims of the present work. We restrict our measurements to  $T \geq 4.6$  K, which is well above the onset of antiferromagnetic order, at  $T_N = 70$  mK (which competes with the Kondo effect), and is within a regime in which the heavy quasiparticles are still well defined, that is, safely away from the quantum critical point<sup>17</sup>.  $\text{YbRh}_2\text{Si}_2$  is particularly well suited for STM investigations because it can be perfectly cleaved (see below) and because it is possible to compare the consequent scanning tunnelling spectroscopy (STS) results with renormalized band-structure calculations<sup>21</sup> even at low energy.

$\text{YbRh}_2\text{Si}_2$  crystallizes in the tetragonal  $\text{ThCr}_2\text{Si}_2$  structure<sup>19</sup> with lattice parameters  $a = 4.007$  Å and  $c = 9.858$  Å (Fig. 1a, inset). A typical topography obtained at  $T = 4.6$  K is presented in Fig. 1 and shows that the crystal has cleaved perpendicular to the crystallographic  $c$  axis. We observe the expected square arrangement of the surface atoms, with a spacing that corresponds to either an ytterbium- or silicon-terminated surface (cleavage of  $\text{YbRh}_2\text{Si}_2$  is expected to take place<sup>22</sup> between the ytterbium and silicon planes). In accord with the low residual resistivity of our samples ( $\rho_0 \approx 0.5$  μΩ cm; see Supplementary Information, section I), we find a very low density of defects. These defects cause changes in height of the order of 20 pm on top of the atomic corrugations of the regular lattice (Fig. 1d). An analysis of these defects, along with other observations outlined below, indicates that Fig. 1 represents a silicon-terminated surface; details of this analysis, as well as an example of an ytterbium-terminated surface, are given in Supplementary Information, section II, and Supplementary Fig. 1, respectively. In the following, we focus on silicon-terminated surfaces, which are (as in pure  $\text{URu}_2\text{Si}_2$  (ref. 6)) more frequently observed.

An overview of our experimental conductance data,  $g(V, T) = dI/dV$ , where  $I$  and  $V$  respectively denote current and voltage, is presented in Fig. 2a. For clarity, only results for a few selected temperatures are shown,

<sup>1</sup>Max Planck Institute for Chemical Physics of Solids, Nöthnitzer Strasse 40, 01187 Dresden, Germany. <sup>2</sup>Max Planck Institute for the Physics of Complex Systems, Nöthnitzer Strasse 38, 01187 Dresden, Germany. <sup>3</sup>Institut für Mathematische Physik, TU Braunschweig, Mendelssohnstrasse 3, 38106 Braunschweig, Germany.

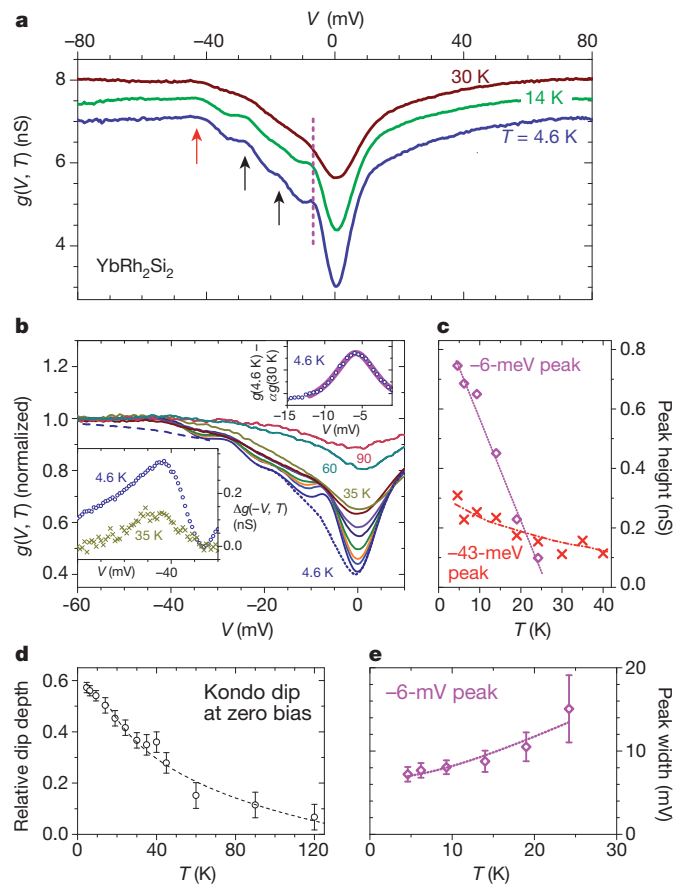


**Figure 1 | Topography of a cleaved  $\text{YbRh}_2\text{Si}_2$  single crystal at 4.6 K.** **a**, View ( $18 \times 18 \text{ nm}^2$ ) exemplifying the almost perfect surface in the  $a$ - $b$  plane of the tetragonal crystal structure shown in the inset. There was no sign of any other structure in any of the six investigated single crystals. Gap voltage, 0.3 V; current set point, 0.6 nA. **b**, Magnified view of **a** ( $2 \times 2 \text{ nm}^2$ ). Shown are raw data except for a plane correction. **c**, Fast Fourier transform of topography in **a**, yielding a distance between the surface atoms of 4 Å. **d**, Magnified scans along the lines through two defects indicated in **a**. Data (blue line) acquired almost perpendicular to the fast scan direction indicate excellent stability.

with data at  $T = 14$  and 30 K offset and a small, linear background contribution subtracted. At  $T = 4.6$  K, we clearly resolve three features with apparently different origins. First,  $g(V, T)$  has a V-shaped depression around zero bias voltage, that is, for  $|V| \lesssim 10$  mV. As shown below, this scale is related<sup>23</sup> to  $T_K$  and we henceforth refer to this feature as the 'Kondo dip'. Second, we observe three peaks at around  $-17$ ,  $-27$  and  $-43$  mV that are only weakly temperature dependent (Fig. 2a, arrows). The two more clearly developed peaks, at  $-27$  and  $-43$  mV, although broadened, have been resolved at temperatures of up to 40 K. Third, a more strongly temperature-dependent peak is located at  $V \approx -6$  mV (dashed line). We discuss these features separately in the following.

The most clear-cut assignment is for the three peaks at  $-17$ ,  $-27$  and  $-43$  mV. These are due to CEF excitations: inelastic neutron scattering measurements<sup>16</sup> revealed such excitations at 17, 25 and 43 meV, in excellent agreement with our STS results. Moreover, they are clearly reflected in the renormalized band-structure calculation<sup>21</sup> (Fig. 3e). The observation of these peaks underlines that bulk properties of  $\text{YbRh}_2\text{Si}_2$  are predominantly probed by our STS. Because the CEF excitations originate in ytterbium, this again makes an ytterbium-terminated surface unlikely and rather points to termination by silicon. We emphasize that CEF excitations in an ytterbium, that is, hole, system with occupied CEF levels are expected at negative voltages in STS (in contrast to cerium-based Kondo systems) and follow naturally from the Kondo picture in multi-orbital systems<sup>24</sup>.

To investigate the temperature evolution of these CEF-related peaks, we focus on the most prominent one, at  $-43$  mV, and make use of the surprising experimental finding of a nearly symmetric



**Figure 2 | Tunnelling spectroscopy of  $\text{YbRh}_2\text{Si}_2$ .** **a**, Overview of tunnelling conductance,  $g(V, T)$ , at selected temperatures. Spectra at  $T = 14$  and 30 K are offset for clarity. **b**, Spectra normalized to  $g(V = -80 \text{ mV}, T)$ . The dotted line represents a background (data at 30 K scaled by a factor  $\alpha$  that accounts for the temperature evolution of the Kondo dip) for  $g(V, T = 4.6 \text{ K})$  on top of which the peak at approximately  $-6$  mV (upper inset) is superposed. The dashed line shows data at 4.6 K 'mirrored' from positive bias. The lower inset shows a magnified view of the peak at  $-43$  mV after subtraction of mirrored positive-bias data. **c–e**, Temperature dependences estimated from **b**. **c**, CEF excitations (arrows in **a**) exemplified for the  $-43$ -meV peak (red) and the Kondo lattice peak at  $-6$  meV (pink; dashed line in **a**). Lines are guides to the eye. **d**, Relative depth,  $h_0$ , of the Kondo dip at zero bias; the line is a logarithmic fit. **e**, Width of the peak at  $-6$  mV; the line is a fit to equation (2). Locally resolved spectroscopy at  $T = 4.6$  K did not indicate any spatial variation of spectroscopic features (Supplementary Information, section III). Our spatial homogeneity is in accord with a silicon-terminated surface. Error bars, s.e. of the fitted parameters.

$g(V, T)$  profile. If the  $g(V, T)$  data measured for positive voltages are mirrored to negative voltages, they appear to form a background on top of which the peak resides (Supplementary Information, section V). This is exemplified for  $g(V, T = 4.6 \text{ K})$  by the dashed line in Fig. 2b. For  $T \lesssim 40$  K, the peak at  $-43$  mV is clearly visible after this background subtraction, as shown in the lower inset of Fig. 2b for two representative temperatures. For all our measured samples, the peak is located without any apparent temperature dependence at  $-43 \pm 2$  mV, an observation that is consistent with its ascription to CEF excitations. The temperature dependence of this peak's height is shown in Fig. 2c.

We now discuss the Kondo dip in our STS spectra around zero bias voltage. These conductance curves (Fig. 2a) are to be compared with those obtained on single magnetic Kondo impurities<sup>9–12</sup>, where tunnelling not only into the conduction band but also into the localized state contributes to the STS signal. The coupling between the two tunnelling channels results in a typical Fano resonance<sup>45</sup> at energy  $E_0$  and of width  $\Gamma$ :  $g(V) \propto (\eta + q)^2/(\eta^2 + 1)$ . Here  $\eta = 2(eV - E_0)/\Gamma$  and the asymmetry parameter,  $q$ , relates the two tunnelling channels.  $\text{YbRh}_2\text{Si}_2$

contains a dense lattice of magnetic ions, and simultaneous tunnelling processes are again possible<sup>25–27</sup>. As a result, the experimentally obtained  $g(V, T)$  cannot be directly related to the local conduction (that is, *spd*) electron density of states (DOS),  $\rho_{RR}(\varepsilon) = (1/\pi)\text{Im}\langle \mathbf{R} | \tilde{G}(\varepsilon) | \mathbf{R} \rangle$ , where  $\tilde{G}(\varepsilon)$  is the Green function for the bare conduction electron and  $\mathbf{R}$  is the sample position closest to the STM tip. For a conceptual understanding of the Kondo dip around zero bias and its temperature dependence, however, it suffices to consider tunnelling into the conduction band as the dominant channel and ignore any direct tunnelling into the ytterbium  $4f$  states (Supplementary Information, section IV). This approximation is justified by our focus on silicon-terminated surfaces. Owing to the Kondo coupling between conduction electrons and the lattice of  $4f$  moments, the STS signal still contains information on the ytterbium  $4f$  DOS near the Fermi energy,  $E_F$ , and the Fano effect occurs. For a single Kondo impurity, this has been explicitly demonstrated previously<sup>28</sup>. For the heavy-fermion lattice, the difference between the experimentally obtained  $g(V, T)$  and  $\rho_{RR}(\varepsilon)$  can be expressed in terms of the difference between the imaginary part of the fully renormalized single-electron Green function,  $\langle \mathbf{R} | G(\varepsilon) | \mathbf{R} \rangle$ , and  $\rho_{RR}(\varepsilon)$ :

$$g(V, T) \propto \frac{1}{\pi} \text{Im} \langle \mathbf{R} | G(\varepsilon) - \tilde{G}(\varepsilon) | \mathbf{R} \rangle + \frac{1}{\pi} \text{Im} \langle \mathbf{R} | \tilde{G}(\varepsilon) | \mathbf{R} \rangle$$

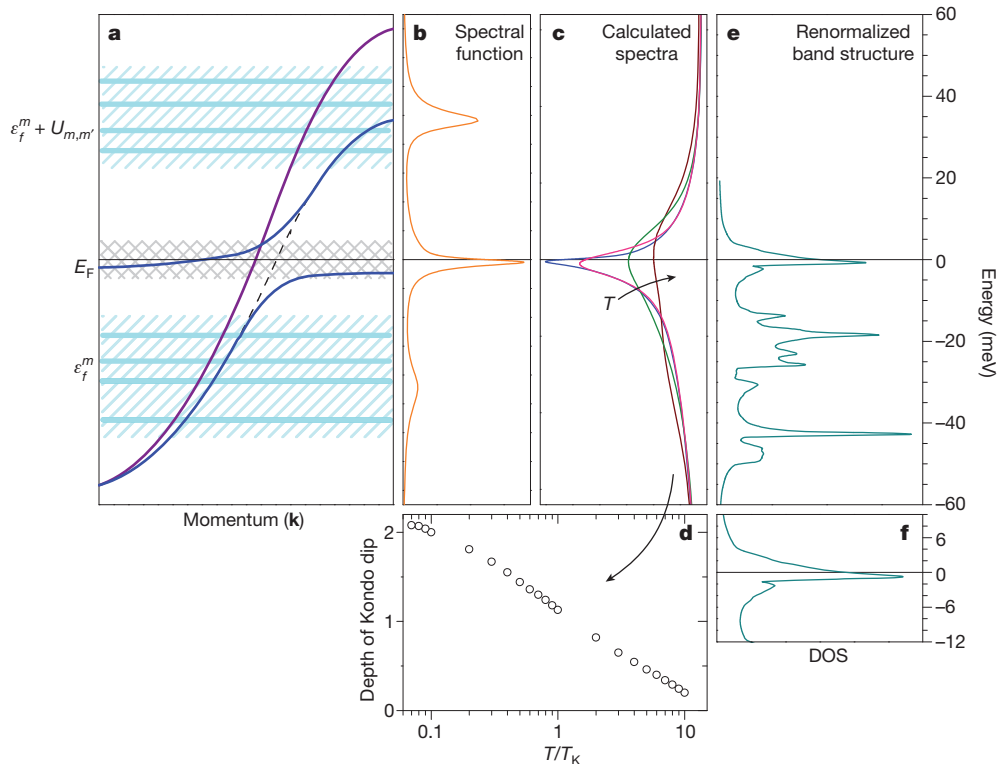
$$\langle \mathbf{R} | G(\varepsilon) - \tilde{G}(\varepsilon) | \mathbf{R} \rangle = G_{RR}(\varepsilon) - \tilde{G}_{RR}(\varepsilon)$$

$$= \sum_{n=1}^{\infty} \sum_{\mathbf{r}_1 \dots \mathbf{r}_n} \tilde{G}_{R\mathbf{r}_1} \tilde{G}_{\mathbf{r}_1 \mathbf{r}_2} \dots \tilde{G}_{\mathbf{r}_n \mathbf{R}} (T(\varepsilon))^n$$
(1)

Here  $\sum_{\mathbf{r}_1 \dots \mathbf{r}_n}$  denotes a sum over all ytterbium sites,  $\mathbf{r}_i$ , of the Kondo

lattice, with each site occurring at most once in such a way that each sequence  $\{\mathbf{R}, \mathbf{r}_i (i = 1, \dots, n), \mathbf{R}\}$  forms a self-avoiding loop.  $T(\varepsilon)$  is the  $T$  matrix and is proportional to the ytterbium  $4f$  DOS (Fig. 3b). We take two complementary approaches to addressing this problem (Supplementary Information, section IV). First, equation (1) can be evaluated approximately by considering only the first term of the sum. This approximation cannot describe the formation of a hybridization gap in a Kondo lattice, yet it yields reasonable results at increased energies and temperatures. The calculated spectra are presented for several temperatures in Fig. 3c, and the resulting temperature evolution of the depth of the Kondo dip is shown in Fig. 3d. A comparison between these results and experimental data of Fig. 2d can be found in Supplementary Information, section IV. Second, a Fermi-liquid-based renormalized band-structure calculation<sup>21</sup> for the lattice-coherent features, for example the incomplete hybridization gap and the full CEF scheme (implemented only incompletely in Fig. 3a), complements the first approach. This is shown in Fig. 3e, f. These considerations offer insight into the nature of the Kondo dip: loosely speaking, the on-site Kondo interaction—that is, the formation of local Kondo singlets—diminishes the conduction electron DOS available for tunnelling, which causes a reduction in conductance below  $|V| \approx 10$  mV, a scale that roughly corresponds to  $T_K$ .

The Kondo dip smoothly becomes shallower with increasing  $T$  (Fig. 2b), albeit substantially faster than is expected for thermal broad-



**Figure 3 | Heavy-quasiparticle formation in a Kondo lattice.** **a**, The renormalized quasiparticle bands<sup>29</sup>. The local ytterbium  $4f$  levels well below  $E_F$  split up into four Kramers doublets,  $\varepsilon_f^m < E_F$ ,  $m = 1, \dots, 4$ . The doubly occupied states are located at  $\varepsilon_f^m + U_{m,m'} < E_F$ ,  $m, m' = 1, \dots, 4$ . Owing to the hybridization between the conduction band and the  $4f$  moments, a hybridization gap occurs near  $E_F$  in the quasiparticle bands at low temperatures. The hole character of the ytterbium  $4f^{13}$  states requires  $E_F$  to be located in the upper quasiparticle band. The red curve shows a band that, owing to symmetry, does not hybridize with the  $4f$  states and remains unaffected. **b**, Local spectral function obtained within the multi-level, finite-Coulomb-repulsion non-crossing approximation (Supplementary Information, section

IV). The ytterbium hole character results in a Kondo resonance slightly below  $E_F$ . **c**, The first term of equation (1) evaluated with the multi-level, finite-Coulomb-repulsion non-crossing approximation at several temperatures (increasing in the direction of the arrow), demonstrating the formation of the Kondo dip. **d**, Depth of the Kondo dip obtained from **c**. **e**, Quasiparticle DOS for YbRh<sub>2</sub>Si<sub>2</sub> within a renormalized band-structure calculation<sup>21</sup>. The renormalized CEF positions,  $\tilde{\varepsilon}_f^m$ , agree with our STS data. We note that the  $\tilde{\varepsilon}_f^m$  are much smaller than the atomic energies,  $\varepsilon_f^m$ . **f**, The region of **e** around  $E_F$ : a small pseudogap has formed at around  $-2$  meV. The formation of the incomplete hybridization gap is discussed in the Supplementary Information, section IV.



ening alone. Its depth,  $h_0$ , is determined as described in Supplementary Information, section V, and  $h_0(T)$  is presented in Fig. 2d. It can be reasonably well described by a logarithmic temperature dependence within a range that is centred on  $T_K = 80\text{--}100\text{ K}$  (Fig. 2d, dashed line). As already mentioned, such behaviour is consistent with the temperature dependence of both the thermopower<sup>20</sup> and the resistivity<sup>19,20</sup> of  $\text{YbRh}_2\text{Si}_2$ , which show maxima at 80 and 100 K, respectively. Whereas the former maximum is a good measure of  $T_K$ , the latter results from a decreasing Kondo scattering of the charge carriers<sup>18</sup> when, on cooling the system, the  $4f$  electrons increasingly ‘condense’ into their CEF-derived Kramers doublet ground state<sup>19,20</sup>. Hence, the observed temperature dependence of the Kondo dip supports our conjecture that this reduced  $g(V)$  for small  $V$  is related to the on-site Kondo interaction between the  $4f$  states and the conduction electrons.

The nearly symmetric voltage dependence of the Kondo dip is compatible with tunnelling into the conduction electron band being predominant<sup>10,26</sup>. This predominance is consistent with a silicon-terminated surface and supports the corresponding approximation of equation (1).

The lowest-energy peak (at  $V \approx -6\text{ mV}$ ; the third feature of the conductance data pointed out above) and, in particular, its temperature dependence can be accounted for neither by CEF excitations nor within a single-ion picture. Instead, both are naturally explained as a generic Kondo lattice effect<sup>29,30</sup>. Moreover, the renormalized band-structure calculation yields a small hybridization gap a few millielectronvolts below  $E_F$  (Fig. 3f). In real systems, a fully developed hybridization gap is often not expected, owing to the momentum ( $\mathbf{k}$ ) dependence of the hybridization and possible magnetic correlations, in contrast to model calculations<sup>25,26</sup>. Our theoretical considerations above indicate that the peak at about  $-6\text{ meV}$  signals the formation of an incomplete hybridization gap (Fig. 3a, f), in accordance with our renormalized band-structure calculations for  $\text{YbRh}_2\text{Si}_2$  and general considerations for a Kondo lattice<sup>29,30</sup>.

An analysis of the peak at  $V \approx -6\text{ mV}$  is severely hampered by the nearby strong Kondo dip onto which this peak is superimposed. The procedure applied to decompose these two features is described in Supplementary Information, section V. The results for the  $g(V, T = 4.6\text{ K})$  data are presented in the upper inset of Fig. 2b. This peak as well as those with  $4.6\text{ K} < T < 30\text{ K}$  can best be fitted by Gaussians (Fig. 2b, line in upper inset), which allows us to extract their heights, positions and widths. The peak height quickly decreases with temperature (Fig. 2c) and is estimated to be completely suppressed at  $T \approx 27\text{ K}$ , in good agreement with an estimate<sup>20</sup> of  $29\text{ K}$  for the Kondo temperature of the CEF ground-state doublet<sup>18</sup> in  $\text{YbRh}_2\text{Si}_2$ . Independently of temperature, the peak remains at  $-6\text{ meV}$ , an energy equivalent to a temperature of the order of, but slightly less than, the single-ion  $T_K$ , as expected for a hybridization gap. Rather than the exact position in energy of the Kondo lattice feature, which in a real material is a complex quantity, we analyse its width,  $w$ . In the single-impurity case, the zero-temperature width depends exponentially on the hybridization and local conduction DOS, reflecting the many-body character of the Kondo resonance. The increased broadening of the feature at finite temperature can, in contrast to the peak height, be accounted for by the relation<sup>31</sup>

$$w = 2\sqrt{(\pi k_B T)^2 + 2(k_B T_K)^2} \quad (2)$$

where  $k_B$  is Boltzmann’s constant. For our case of a coherent Kondo lattice, a similar exponential dependence (with modified parameters) can be expected. By analogy to equation (2), we analyse our experimentally obtained  $w(T)$  in terms of an entirely phenomenological expression:  $w = 2\sqrt{(\pi k_B T)^2 + 2(k_B T_L)^2}$ . A fit (Fig. 2e, dashed line) yielded  $T_L = 30 \pm 6\text{ K}$  for the zero-temperature width of the lattice-coherent Kondo feature. We note that our analysis of  $w(T)$  does not require the hybridization gap to be fully developed.

These results give insight into the thermal evolution of the Kondo effect in  $\text{YbRh}_2\text{Si}_2$ . The emerging local Kondo effect, which develops at  $T_K = 80\text{--}100\text{ K}$ , averages over all involved CEF levels<sup>18</sup>. On cooling,

excited CEF states become depopulated, and below  $T_L \approx 30\text{ K}$  only the lowest-lying CEF Kramers doublet is occupied. This allows for the development of a spatially coherent state that is manifested by an additional peak in  $g(V, T)$ , at  $-6\text{ mV}$ , reflecting a ‘Kondo lattice resonance’ related to an incomplete hybridization gap forming at around the single-ion Kondo temperature of the lowest-lying CEF doublet. Thus, by investigating the low-energy charge excitation spectrum we are able to illustrate the progressive Kondo screening and the emergence of coherent heavy-fermion states. This spectroscopic approach opens the way to a detailed microscopic study of the regime at even lower temperatures and will shed new light on, for example, quantum critical electronic excitations.

## METHODS SUMMARY

We grew single crystals of  $\text{YbRh}_2\text{Si}_2$  by an indium-flux method. The platelet-shaped samples (thin dimension parallel the crystallographic  $c$  direction) were cleaved *in situ* under ultrahigh-vacuum conditions, at low temperature ( $\sim 20\text{ K}$ ) and parallel to the crystallographic  $a$ – $b$  plane. STS was conducted after stabilizing each measurement temperature for several hours and by using a standard a.c. lock-in amplifier technique (typical bias modulation of  $200\text{ }\mu\text{V}$ , averaging for  $200\text{ ms}$  at each measured bias voltage). We used a generalized multi-level non-crossing approximation to account roughly for the CEF splitting of the ytterbium  $4f^{13}$  state in the calculation of the current–voltage characteristics. Lattice coherence effects were compared with renormalized band-structure calculations for  $\text{YbRh}_2\text{Si}_2$ .

Received 14 February; accepted 21 April 2011.

- Shiba, S. & Kuramoto, Y. (eds) Kondo effect – 40 years after the discovery. *J. Phys. Soc. Jpn* **74** (special topic), 1–238 (2005).
- Burdin, S., Georges, A. & Grepel, D. R. Coherence scale of the Kondo lattice. *Phys. Rev. Lett.* **85**, 1048–1051 (2000).
- Yang, Y., Fisk, Z., Lee, H.-O., Thompson, J. D. & Pines, D. Scaling the Kondo lattice. *Nature* **454**, 611–613 (2008).
- Fano, U. Effects of configuration interaction on intensities and phase shifts. *Phys. Rev.* **124**, 1866–1878 (1961).
- Schiller, A. & Hershfield, S. Theory of scanning tunneling spectroscopy of a magnetic adatom on a metallic surface. *Phys. Rev. B* **61**, 9036–9046 (2000).
- Schmidt, A. R. *et al.* Imaging the Fano lattice to ‘hidden order’ transition in  $\text{URu}_2\text{Si}_2$ . *Nature* **465**, 570–576 (2010).
- Aynajian, P. *et al.* Visualizing the formation of the Kondo lattice and the hidden order in  $\text{URu}_2\text{Si}_2$ . *Proc. Natl Acad. Sci. USA* **107**, 10383–10388 (2010).
- Coleman, P. & Schofield, A. J. Quantum criticality. *Nature* **433**, 226–229 (2005).
- Madhavan, V., Chen, W., Jamneala, T., Crommie, M. F. & Wingreen, N. S. Tunneling into a single magnetic atom: spectroscopic evidence of the Kondo resonance. *Science* **280**, 567–569 (1998).
- Li, J., Schneider, W.-D., Berndt, R. & Delley, B. Kondo scattering observed at a single magnetic impurity. *Phys. Rev. Lett.* **80**, 2893–2896 (1998).
- Wahl, P. *et al.* Kondo temperature of magnetic impurities at surfaces. *Phys. Rev. Lett.* **93**, 176603 (2004).
- Hirjibehedin, C. F. *et al.* Large magnetic anisotropy of a single atomic spin embedded in a surface molecular network. *Science* **317**, 1199–1203 (2007).
- Prüser, H. *et al.* Long-range Kondo signature of a single magnetic impurity. *Nature Phys.* **7**, 203–206 (2011).
- Cox, D. L. & Zawadowski, A. Exotic Kondo effects in metals: magnetic ions in a crystalline electric field and tunnelling centres. *Adv. Phys.* **47**, 599–942 (1998).
- Fujimori, S.-I. *et al.* Electronic structure of heavy fermion uranium compounds studied by core-level photoelectron spectroscopy. *J. Phys. Soc. Jpn.* (in the press).
- Stockert, O. *et al.* Crystalline electric field excitations of the non-Fermi-liquid  $\text{YbRh}_2\text{Si}_2$ . *Physica B* **378**, 157–158 (2006).
- Custers, J. *et al.* The break-up of heavy electrons at a quantum critical point. *Nature* **424**, 524–527 (2003).
- Cornut, B. & Coqblin, B. Influence of the crystalline field on the Kondo effect of alloys and compounds with cerium impurities. *Phys. Rev. B* **5**, 4541–4561 (1972).
- Trovarelli, O. *et al.*  $\text{YbRh}_2\text{Si}_2$ : pronounced non-Fermi-liquid effects above a low-lying magnetic phase transition. *Phys. Rev. Lett.* **85**, 626–629 (2000).
- Köhler, U., Oeschler, N., Steglich, F., Maquilon, S. & Fisk, Z. Energy scales of  $\text{Lu}_{1-x}\text{Yb}_x\text{Rh}_2\text{Si}_2$  by means of thermopower investigations. *Phys. Rev. B* **77**, 104412 (2008).
- Friedemann, S. *et al.* Hall effect measurements and electronic structure calculations on  $\text{YbRh}_2\text{Si}_2$  and its reference compounds  $\text{LuRh}_2\text{Si}_2$  and  $\text{YbIr}_2\text{Si}_2$ . *Phys. Rev. B* **82**, 035103 (2010).
- Danzonbächer, S. *et al.* Momentum dependence of  $4f$  hybridization in heavy-fermion compounds: angle-resolved photoemission study of  $\text{YbIr}_2\text{Si}_2$  and  $\text{YbRh}_2\text{Si}_2$ . *Phys. Rev. B* **75**, 045109 (2007).
- Costi, T. A. & Manini, N. Low-energy scales and temperature-dependent photoemission of heavy fermions. *J. Low-Temp. Phys.* **126**, 835–866 (2002).
- Kroha, J. *et al.* Structure and transport in multi-orbital Kondo systems. *Physica E* **18**, 69–72 (2003).
- Mal'tseva, M., Dzero, M. & Coleman, P. Electron cotunneling into a Kondo lattice. *Phys. Rev. Lett.* **103**, 206402 (2009).

26. Figgins, J. & Morr, D. K. Differential conductance and quantum interference in Kondo systems. *Phys. Rev. Lett.* **104**, 187202 (2010).
27. Wölfle, P., Dubi, Y. & Balatsky, A. V. Tunneling into clean heavy fermion compounds: origin of the Fano line shape. *Phys. Rev. Lett.* **105**, 246401 (2010).
28. Újsághy, O., Kroha, J., Szunyogh, L. & Zawadowski, A. Theory of the Fano resonance in the STM tunneling density of states due to a single Kondo impurity. *Phys. Rev. Lett.* **85**, 2557–2560 (2000).
29. Martin, R. M. Fermi-surface sum rule and its consequences for periodic Kondo and mixed-valence systems. *Phys. Rev. Lett.* **48**, 362–365 (1982).
30. Kim, C.-I., Kuramoto, Y. & Kasuya, T. Self-consistent dynamical theory for the Anderson lattice. *J. Phys. Soc. Jpn* **59**, 2414–2425 (1990).
31. Nagaoka, K., Jamneala, T., Grobis, M. & Crommie, M. F. Temperature dependence of a single Kondo impurity. *Phys. Rev. Lett.* **88**, 077205 (2002).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank P. Coleman, T. Costi, A. Georges, D. Morr, J. Schmalian, S. Seiro, Q. Si and P. Wölfle for discussions. We are indebted to J. C. Davis for comments on our manuscript. This work is partly supported by the German Research Foundation through DFG Forschergruppe 960.

**Author Contributions** F.S. and S.W. designed the project. S.E. performed measurements, S.E. and S.W. conducted the data analysis. S.K. provided the theoretical framework and non-crossing approximation calculations. G.Z. did the band-structure calculations. C.K. and C.G. synthesized and characterized the materials. S.W., S.K. and F.S. wrote the manuscript.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at [www.nature.com/nature](http://www.nature.com/nature). Correspondence and requests for materials should be addressed to S.W. ([wirth@cpfs.mpg.de](mailto:wirth@cpfs.mpg.de)).

# Modular and predictable assembly of porous organic molecular crystals

James T. A. Jones<sup>1</sup>, Tom Hasell<sup>1</sup>, Xiaofeng Wu<sup>1</sup>, John Bacsá<sup>1</sup>, Kim E. Jelfs<sup>1</sup>, Marc Schmidtman<sup>1</sup>, Samantha Y. Chong<sup>1</sup>, Dave J. Adams<sup>1</sup>, Abbie Trewin<sup>1</sup>, Florian Schiffman<sup>2</sup>, Furio Cora<sup>2</sup>, Ben Slater<sup>2</sup>, Alexander Steiner<sup>1</sup>, Graeme M. Day<sup>3</sup> & Andrew I. Cooper<sup>1</sup>

Nanoporous molecular frameworks<sup>1–7</sup> are important in applications such as separation, storage and catalysis. Empirical rules exist for their assembly but it is still challenging to place and segregate functionality in three-dimensional porous solids in a predictable way. Indeed, recent studies of mixed crystalline frameworks suggest a preference for the statistical distribution of functionalities throughout the pores<sup>7</sup> rather than, for example, the functional group localization found in the reactive sites of enzymes<sup>8</sup>. This is a potential limitation for ‘one-pot’ chemical syntheses of porous frameworks from simple starting materials. An alternative strategy is to prepare porous solids from synthetically preorganized molecular pores<sup>9–15</sup>. In principle, functional organic pore modules could be covalently prefabricated and then assembled to produce materials with specific properties. However, this vision of mix-and-match assembly is far from being realized, not least because of the challenge in reliably predicting three-dimensional structures for molecular crystals, which lack the strong directional bonding found in networks. Here we show that highly porous crystalline solids can be produced by mixing different organic cage modules that self-assemble by means of chiral recognition. The structures of the resulting materials can be predicted computationally<sup>16,17</sup>, allowing *in silico* materials design strategies<sup>18</sup>. The constituent pore modules are synthesized in high yields on gram scales in a one-step reaction. Assembly of the porous co-crystals is as simple as combining the modules in solution and removing the solvent. In some cases, the chiral recognition between modules can be exploited to produce porous organic nanoparticles. We show that the method is valid for four different cage modules and can in principle be generalized in a computationally predictable manner based on a lock-and-key assembly between modules.

A basic tool in the synthesis of functional extended solids is the ability to combine different chemical entities in a controlled and modular fashion. This has been demonstrated for structurally related, or ‘isorecticular’, porous metal–organic frameworks (MOFs)<sup>3</sup>. MOFs can be prepared with more than one chemical function, either by direct reaction of mixed precursors<sup>7</sup> or by post-synthetic modification<sup>19</sup>. Although both MOFs and zeolites can comprise fused, compartmentalized cages, it is still generally challenging to segregate structural units in a programmed and predictable way.

Most nanoporous networks are synthesized in ‘one-pot’ chemical reactions where all of the precursors are mixed together simultaneously<sup>1–7</sup>. The three-dimensional network structure arises from self-assembly of the components. By contrast, natural products are synthesized in stepwise reaction sequences where isolable molecular intermediates are elaborated and combined to create more complex structures. An analogous, supramolecular strategy<sup>20</sup> for porous organic solids would be to preorganize larger chemical subunits, or pore modules, before assembling the extended crystal. This approach requires building blocks, or tectons<sup>21,22</sup>, that are self-assembling,

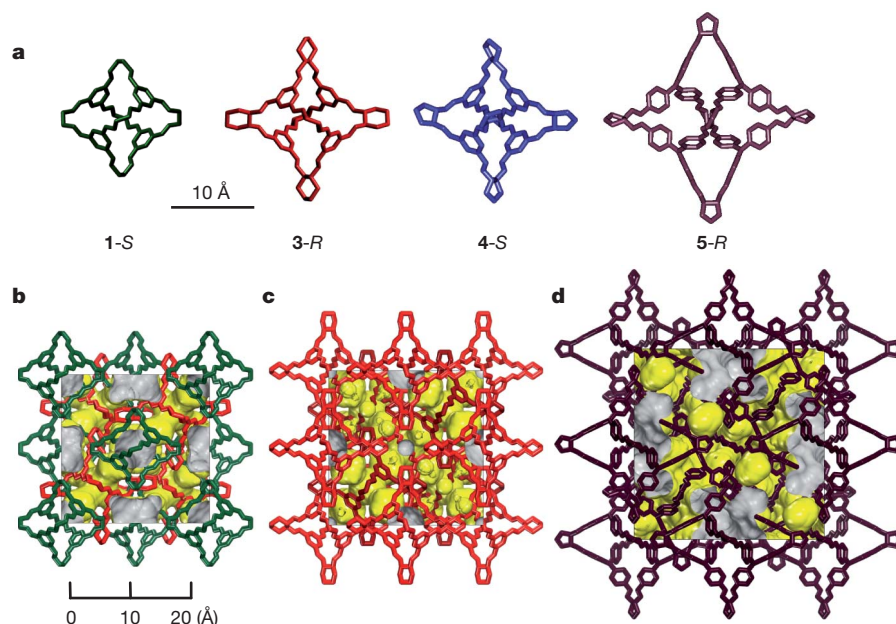
prefabricated molecular analogues of the secondary building units in networks such as MOFs<sup>1–5,7</sup>. To be broadly useful, the pore modules should pack together in predictable ways. Individual modules could then be designed to incorporate desirable chemical functionalities, either by chemical derivatization<sup>15</sup> or by physical encapsulation within the molecular pores<sup>23</sup>. Mixing different functional modules might produce porous solids with unusual properties, perhaps, for example, by combining both acid- and base-containing cage modules within the same porous solid along with vacant, flow-through pores. In practice, however, many components of this strategy are currently missing. Although a large number of porous molecular solids are known<sup>9–15</sup>, as highlighted in a recent review<sup>24</sup>, the rules that underpin their three-dimensional, non-covalent assembly are poorly understood. In this respect, the notion of ‘supramolecular synthesis’<sup>20</sup> is still unfulfilled. Levels of porosity in such molecular organic solids are also modest: until recently<sup>12–15</sup>, Brunauer–Emmett–Teller specific surface areas,  $S_{\text{ABET}}$ , of less than  $400 \text{ m}^2 \text{ g}^{-1}$  were typical<sup>24</sup>. Moreover, porous molecular solids could not be described as modular because almost all examples are single-component crystals.

In this study, we report the production of porous organic molecular co-crystals, thus demonstrating a new modular assembly concept. We also describe computational methods to predict these crystal structures *ab initio*, greatly enhancing the long-term prospects for rational materials design<sup>18</sup>. The materials were fabricated from combinations of the four pore modules shown in Fig. 1a. The first porous co-crystal was constructed from two organic cages that we described previously<sup>12</sup>: cage 1 and cage 3-*R*. Porosity is covalently prefabricated in the individual tetrahedral cage molecules such that each module has four triangular pore windows with diameters of around 6 Å (Fig. 1a; see also scheme 1 in Supplementary Information). Each cage is just over 1 nm in size. Both cage modules have helical chirality: 1 comprises, in crystalline form, an equimolar mixture of the helical enantiomers 1-*S* and 1-*R*<sup>12,14</sup>, whereas 3-*R* is homochiral. Both cages are soluble in common solvents and can be simply mixed together in solution. Slow evaporation of an equimolar solution of 1 and 3-*R* did not lead to separate crystals of the individual modules, but rather to a new single-phase crystalline material. Remarkably, the material is a quasisracemic co-crystal<sup>25</sup>, (1-*S*, 3-*R*). That is, it consists exclusively of the *S* helical enantiomer of 1 crystallized with 3-*R* (Fig. 1b). The apparent loss of the 1-*R* enantiomer, despite 100% sample mass recovery from crystallization, is explained by variable-temperature <sup>1</sup>H NMR measurements. This shows that the helical configurations of 1 interconvert rapidly in solution<sup>14</sup> (Supplementary Fig. 2). The chirality of 1 is therefore dynamically resolved on crystallization with the homochiral cage, 3-*R* (Fig. 2a). Cage 1 is an ‘amphichiral’ module: it can also pair with 3-*S* to form the opposite quasisracemic co-crystal, (1-*R*, 3-*S*). As discussed below, however, this assembly strategy is not limited to dynamically chiral molecules.

The crystal packing for (1-*S*, 3-*R*) is also shown in Fig. 2: the 1-*S* and 3-*R* modules alternate in the crystal lattice in a face-centred cubic

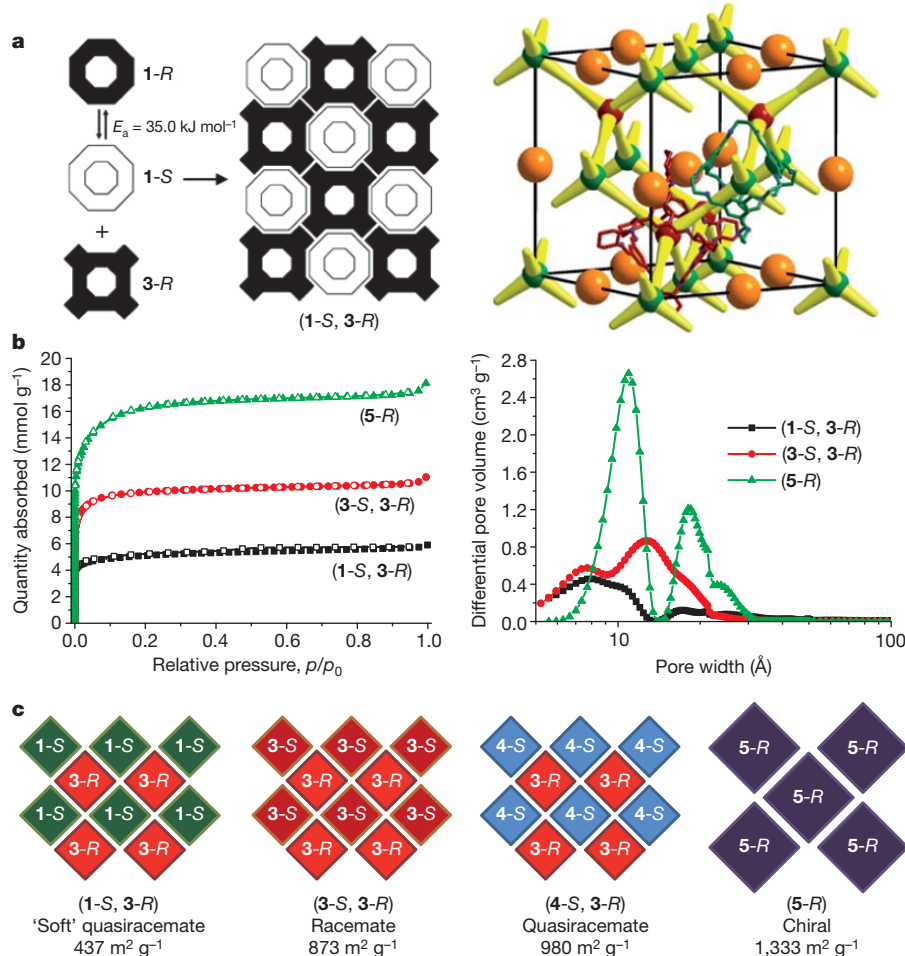
<sup>1</sup>Department of Chemistry and Centre for Materials Discovery, University of Liverpool, Crown Street, Liverpool L69 7ZD, UK. <sup>2</sup>Department of Chemistry, University College London, 20 Gordon Street, London WC1H 0AJ, UK. <sup>3</sup>Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge CB2 1EW, UK.





**Figure 1 | Modular assembly of porous organic cages.** **a**, Structures of four organic cage modules (hydrogen atoms omitted for clarity). Cage 1 is shown as the *S* enantiomer but this module is amphichiral and can interconvert between the *R* and *S*

forms. **b–d**, Crystal structures of porous organic solids formed from these modules, with the Connolly surface shown in yellow (probe radius, 1.82 Å): quasiracemic co-crystal (1-*S*, 3-*R*) (**b**), racemic crystal (3-*S*, 3-*R*) (**c**) and chiral crystal 5-*R* (**d**).



**Figure 2 | Window-to-window assembly results in porosity.** **a**, Helical chirality in 1 is dynamically resolved by heterochiral co-crystallization with 3-*R*. The schematic packing diagram for (1-*S*, 3-*R*) shows the centres of modules 1-*S* and 3-*R* as green and red spheres, respectively; orange spheres represent interstitial voids that are not connected to the diamondoid pore network, which is illustrated in yellow.  $E_a$  is the activation energy for conversion

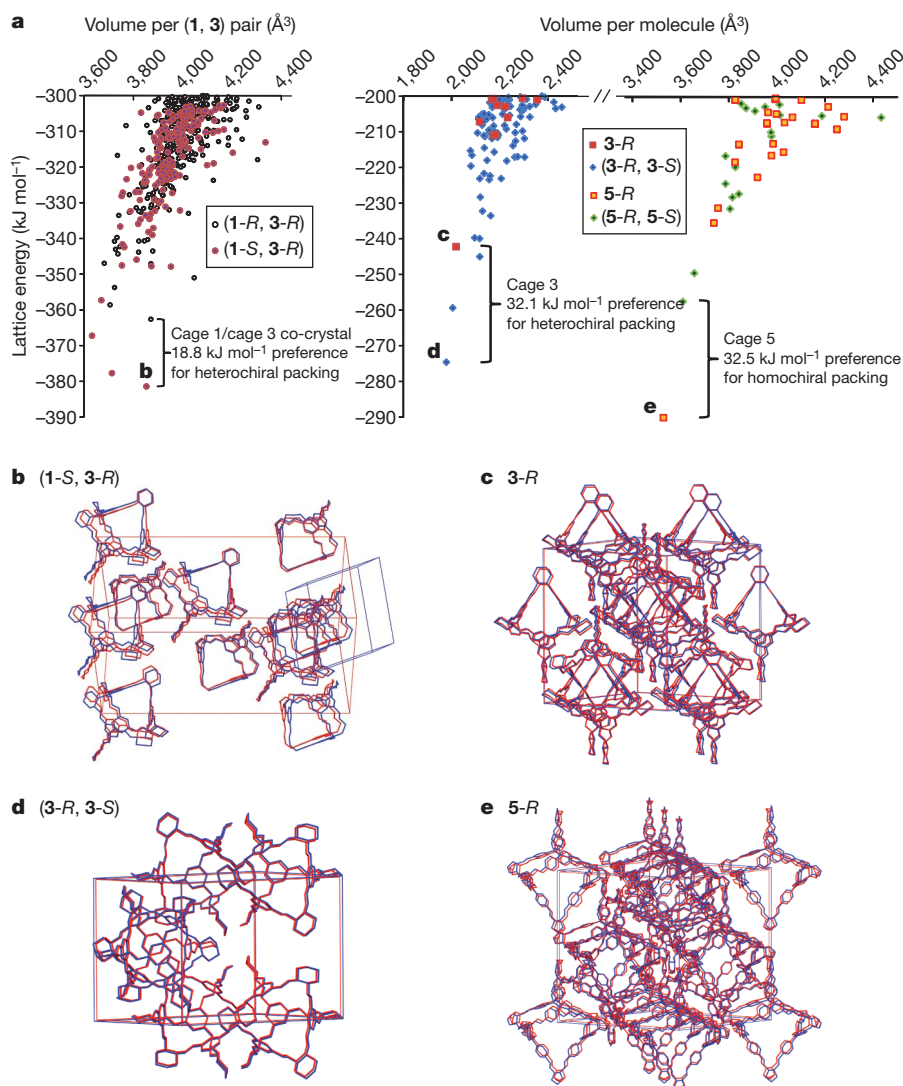
between 1-*S* and 1-*R*, as measured by variable-temperature NMR **b**, Nitrogen gas sorption analysis for crystals and co-crystals shows that pore volume and pore size can be varied systematically, as in isorecticular networks. Filled and open symbols represent sorption and desorption isotherms, respectively.  $p_0$ , atmospheric pressure. **c**, Scheme showing packing for various crystals and co-crystals.

arrangement, analogous to the ZnS ‘zinc blende’ structure. Each cage forms window-to-window interactions with four partner cages of the other type. The result is an interconnected diamondoid pore network. No polymorphs of pure **1** have been found that pack in a window-to-window fashion<sup>12,14</sup>. Therefore, this packing mode is directed by the presence of the chiral co-module, **3-R**. The window-to-window packing arrangement creates permanent micropore channels in the co-crystal, which has a type-I nitrogen sorption isotherm at 77 K (Fig. 2) and a specific surface area of  $S_{\text{BET}} = 437 \text{ m}^2 \text{ g}^{-1}$ . Like the other materials described here, the co-crystal is stable to desolvation and has good thermal stability, showing little weight loss until the onset of decomposition at 350 °C (Supplementary Fig. 7).

The heterochiral pairing (**1-S**, **3-R**) can be considered a directional tecton<sup>22</sup>, comparable to reversible supramolecular interactions such as hydrogen bonding<sup>21</sup> and the ‘sextuple aryl embrace’<sup>26</sup> that involves interlocking aryl rings. Density functional theory (DFT) calculations for isolated cage pairs indicate that the heterochiral window-to-window interaction is  $18 \text{ kJ mol}^{-1}$  more stable than the equivalent homochiral interaction and much more stable than other hypothetical window-to-arene or arene-to-arene pairs that would lead to disconnected pores (Supplementary Fig. 8). Lattice energy calculations confirm that this heterochiral pairing preference carries over to the solid

state and, more significantly, that the observed co-crystal structure can be predicted *ab initio* from the molecular formulae of the modules. Calculations using Monte Carlo simulated annealing to generate hypothetical (**1-S**, **3-R**) crystal structures, followed by energy minimization using anisotropic atom–atom potentials<sup>27,28</sup>, showed the observed packing mode for (**1-S**, **3-R**) to be the global lattice energy minimum (Fig. 3), with good agreement between the *ab initio* predicted structure and the experimental single-crystal X-ray structure (Fig. 3b). The most stable hypothetical homochiral (**1-R**, **3-R**) structure, which lacks window-to-window packing, was predicted to be  $18.8 \text{ kJ mol}^{-1}$  less stable than the observed quasiracemate, (**1-S**, **3-R**). These calculations therefore rationalize the preference for **1** to adopt the **1-S** configuration in the co-crystal and to pack in a window-to-window fashion: that is, both the preferred chirality and the resultant porosity in the solid can be predicted *ab initio*. To verify the atom–atom-potential lattice energy calculations, we performed solid-state DFT calculations on the observed quasiracemate and low-energy predicted homochiral structures: these calculations confirmed the preference for heterochiral packing.

This behaviour is not limited to the pairing of **1-S** and **3-R**. The enantiomers **3-S** and **3-R** also strongly prefer heterochiral window-to-window pairs and assemble in that fashion in a (**3-S**, **3-R**) racemic



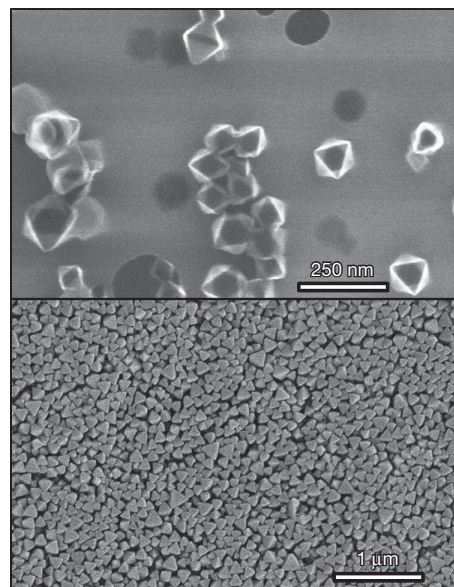
**Figure 3 | Three-dimensional cage assembly can be predicted computationally.** **a**, Lattice energy rankings rationalize the heterochiral packing preference for the (**1**, **3**) co-crystal (structure **b** is favoured over all hypothetical homochiral predicted structures), the racemic packing preference for cage **3** (structure **d** is favoured over **c**) and the chiral preference for

**5** (structure **e** is favoured over all hypothetical racemates). **b–e**, Packing diagrams show the excellent fit between the calculated global-minimum structures (blue) and the experimentally determined structures (red). The predicted (**1-S**, **3-R**) structure in **b** is slightly less symmetrical than the observed **R3** space-group symmetry. The **P1** unit cell is shown.

crystal (Fig. 1c) to give a porous solid with  $SA_{\text{BET}} = 873 \text{ m}^2 \text{ g}^{-1}$ . In this case, the chirality in both modules is fixed rather than dynamic. As before, DFT simulations suggest a significant energy gain ( $19 \text{ kJ mol}^{-1}$ ) in the formation of heterochiral dimers. Again, the crystal structure can be predicted *ab initio*. The experimentally observed racemic packing is the global energy minimum in the set of predicted crystal structures, and there is close agreement between the predicted and observed structures (Fig. 3d). These calculations also suggest a global preference for heterochiral packing modes rather than homochiral. A large energetic gain, of  $32 \text{ kJ mol}^{-1}$ , is calculated for the (3-S, 3-R) racemic crystal over the most stable predicted homochiral structure for 3-R. The global-minimum homochiral prediction also closely reproduces the observed structure for 3-R<sup>12</sup> (Fig. 3c), which, unlike **1**, can be obtained from enantiopure solutions because 3-R does not interconvert with its enantiomer. As for the (1-S, 3-R) co-crystal, the atom–atom lattice energy calculations were verified using periodic DFT calculations, which resulted in similar calculated energy differences (Supplementary Table 1 and Supplementary Fig. 13). An analogous set of experimental observations and crystal structure predictions was obtained for a new cage module, 4-S, which has cyclopentane vertices rather than cyclohexane (Fig. 1a). This module forms a quasiracemic co-crystal, (4-S, 3-R), with  $SA_{\text{BET}} = 980 \text{ m}^2 \text{ g}^{-1}$ . In this case, the predicted global-energy-minimum crystal structure is an ordered version of the most probable site-disordered space-group-*F4*/32 structure, according to powder X-ray data (Supplementary Figs 12 and 15). By itself, 4-S does not pack in a window-to-window fashion (Supplementary Fig. 18). Hence, like (1-S, 3-R), this packing mode is directed by the partner module, 3-R.

Not all systems favour heterochiral assembly and this, too, is predictable from the calculated crystal energy landscape. A new module, 5-R (Fig. 1a), was synthesized by the [4 + 6] cycloaddition reaction between tri(4-formylphenyl)amine and the chiral diamine (*R,R*)-1,2-cyclopentanedi-amine. Cage 5-R is substantially larger than modules **1**, **3** and **4**. For example, the tetrahedron inscribed by the centres of the triangular faces of 5-R has a volume that is 3.8 times larger than the comparable tetrahedron for cage **1** (Supplementary Fig. 24). In this case, lattice energy calculations suggest homochiral window-to-window packing as the clear global energy minimum, and this predicted structure is observed experimentally for 5-R (Fig. 3e); again, DFT calculations agree broadly with the energy differences obtained by atom–atom-potential lattice energy calculations. To our knowledge, 5-R ( $1,702 \text{ g mol}^{-1}$ ) is the largest organic molecule to be successfully tackled by crystal structure prediction<sup>16</sup>. Numerous experiments involving crystallization from mixtures of the modules 5-R and 5-S all led exclusively to homochiral crystals, in agreement with the predicted lattice energy preference over all hypothetical racemic structures. The crystalline solid 5-R has larger pores (compare Fig. 1b, c with Fig. 1d) and a greater pore volume ( $0.63 \text{ cm}^3 \text{ g}^{-1}$ ) than any of the materials produced from the smaller cages **1**, **3** and **4**. The surface area of 5-R ( $SA_{\text{BET}} = 1333 \text{ m}^2 \text{ g}^{-1}$ ) exceeds all but one<sup>15</sup> of the porous molecular (non-network) crystals reported so far<sup>9–14,24</sup> and is comparable with the first generation of covalent organic frameworks<sup>6</sup>. This larger cage shows that it is possible to prepare molecular organic crystals with bespoke pore sizes, analogous to the well-known series of isoreticular MOFs<sup>3</sup> where pore size is defined by organic strut length. A future challenge will be to generalize this non-covalent assembly methodology. Non-identical molecules do not, as a general rule, co-crystallize, and it may be necessary to incorporate specific complementary functionality to induce co-crystallization of dissimilar modules.

We have shown that porous cages can assemble in a modular fashion and, moreover, that the mode of assembly can be predicted accurately using lattice energy calculations. These particular structures are amenable to computation because the directional interlocking of neighbouring cages leads to large energy differences between hypothetical structures. By contrast, most other organic molecules give rise to many



**Figure 4 | Module assembly in solution can be used to produce porous nanoparticles.** Mixing solutions of 3-S and 3-R leads to rapid precipitation of racemic octahedral nanocrystals of (3-S, 3-R) with an average diameter of 130 nm ( $SA_{\text{BET}} = 873 \text{ m}^2 \text{ g}^{-1}$ ). The micrographs show the same sample imaged at two different magnifications.

distinct possible crystal structures that differ in energy by only a few kilojoules per mole<sup>16,17</sup>. Larger, conformationally flexible cage modules would be more challenging for these prediction methods, but significant recent advances have been made in dealing with molecular flexibility<sup>29,30</sup>. Thus, the work presented here opens the way for *in silico* prediction of structure and properties for new candidate porous materials based solely on two-dimensional chemical sketches, thus allowing ‘design by computational selection’.

The solution processability of the cage modules also means that the assembly approach can be extended to achieve structural control beyond the molecular length scale. For example, the (3-S, 3-R) racemate is at least ten times less soluble than the homochiral modules, 3-S and 3-R, and this leads to spontaneous precipitation on mixing of solutions of the two enantiomers (Supplementary Movie 1 and Supplementary Fig. 27). Well-defined, porous (3-S, 3-R) nanocrystals are formed (Fig. 4), thereby translating intermolecular heterochiral tecton interactions into nanoscale morphology control. Porous nanocrystals might make particular applications of these solids possible in future, for example in chiral catalysis or separations.

## METHODS SUMMARY

**Synthesis of compounds.** Cage **1**, cage 4-R and cage 3-R were synthesized in a [4 + 6] cycloaddition reaction involving triformylbenzene and the diamines ethylenediamine, (1S,2S)-cyclopentanedi-amine and (1R,2R)-cyclohexanedi-amine, respectively, using an improved synthetic procedure which produces higher yields than that reported previously<sup>12</sup> (Supplementary Information). Cage 5-R was synthesized by the [4 + 6] cycloaddition reaction between tri(4-formylphenyl)amine and (*R,R*)-1,2-cyclopentanedi-amine. Co-crystals were grown from equimolar solutions of the partner cage modules. Details of the crystallographic analysis, crystal data and gas sorption analysis are described in Supplementary Information.

**Crystal structure prediction.** Crystal structures were generated in the most commonly observed space groups using a Monte Carlo simulated annealing search method. The lowest-energy structures from the Monte Carlo search were then lattice-energy-minimized using anisotropic atom–atom potentials within the crystal structure modelling software DMACRYS<sup>28</sup>. Molecular geometries, generated by DFT single-molecule optimization, were treated as rigid throughout the predictions. Further details are given in Supplementary Information.

**DFT calculations.** Cage pairs and crystal structures were fully optimized in the mixed Gaussian and plane-wave code CP2K<sup>31</sup>, using the TZVP-MOLOPT basis set in combination with Geodecker–Teter–Hutter pseudopotentials and a plane-wave



cut-off of 400 Ry. Molecular and solid-state calculations used the BLYP and PBE functionals, respectively, both with Grimme's D3 dispersion correction<sup>32</sup>.

Received 22 October 2010; accepted 18 April 2011.

- Kondo, M., Yoshitomi, T., Seki, K., Matsuzaka, H. & Kitagawa, S. Three-dimensional framework with channeling cavities for small molecules:  $\{M_2(4,4'-bipy)_3(NO_3)_4 \cdot xH_2O\}_n$  ( $M = Co, Ni, Zn$ ). *Angew. Chem. Int. Edn Engl.* **36**, 1725–1727 (1997).
- Li, H., Eddaoudi, M., O'Keeffe, M. & Yaghi, O. M. Design and synthesis of an exceptionally stable and highly porous metal-organic framework. *Nature* **402**, 276–279 (1999).
- Eddaoudi, M. *et al.* Systematic design of pore size and functionality in isorecticular MOFs and their application in methane storage. *Science* **295**, 469–472 (2002).
- Zhao, X. B. *et al.* Hysteretic adsorption and desorption of hydrogen by nanoporous metal-organic frameworks. *Science* **306**, 1012–1015 (2004).
- Férey, G. *et al.* A chromium terephthalate-based solid with unusually large pore volumes and surface area. *Science* **309**, 2040–2042 (2005).
- Côté, A. P. *et al.* Porous, crystalline, covalent organic frameworks. *Science* **310**, 1166–1170 (2005).
- Deng, H. X. *et al.* Multiple functional groups of varying ratios in metal-organic frameworks. *Science* **327**, 846–850 (2010).
- Nureki, O. *et al.* Enzyme structure with two catalytic sites for double-sieve selection of substrate. *Science* **280**, 578–582 (1998).
- Barbour, L. J. Crystal porosity and the burden of proof. *Chem. Commun.* 1163–1168 (2006).
- Atwood, J. L., Barbour, L. J. & Jerga, A. Storage of methane and freon by interstitial van der Waals confinement. *Science* **296**, 2367–2369 (2002).
- Tranchemontagne, D. J. L., Ni, Z., O'Keeffe, M. & Yaghi, O. M. Reticular chemistry of metal-organic polyhedra. *Angew. Chem. Int. Ed.* **47**, 5136–5147 (2008).
- Tozawa, T. *et al.* Porous organic cages. *Nature Mater.* **8**, 973–978 (2009).
- Bezzu, C. G., Helliwell, M., Warren, J. E., Allan, D. R. & McKeown, N. B. Heme-like coordination chemistry within nanoporous molecular crystals. *Science* **327**, 1627–1630 (2010).
- Jones, J. T. A. *et al.* On-off porosity switching in a molecular organic solid. *Angew. Chem. Int. Ed.* **50**, 749–753 (2011).
- Mastalerz, M., Schneider, M. W., Oppel, I. M. & Presly, O. A salicylbisimine cage compound with high surface area and selective  $CO_2/CH_4$  adsorption. *Angew. Chem. Int. Ed.* **50**, 1046–1055 (2011).
- Day, G. M. *et al.* Significant progress in predicting the crystal structures of small organic molecules – a report on the fourth blind test. *Acta Crystallogr. B* **65**, 107–125 (2009).
- Price, S. L. Computed crystal energy landscapes for understanding and predicting organic crystal structures and polymorphism. *Acc. Chem. Res.* **42**, 117–126 (2009).
- Jansen, M., Doll, K. & Schön, J. C. Addressing chemical diversity by employing the energy landscape concept. *Acta Crystallogr. A* **66**, 518–534 (2010).
- Ingleton, M. J., Barrio, J. P., Guilbaud, J. B., Khimyak, Y. Z. & Rosseinsky, M. J. Framework functionalisation triggers metal complex binding. *Chem. Commun.* 2680–2682 (2008).
- Desiraju, G. R. Supramolecular synthons in crystal engineering – a new organic-synthesis. *Angew. Chem. Int. Edn Engl.* **34**, 2311–2327 (1995).
- Simard, M., Su, D. & Wuest, J. D. Use of hydrogen bonds to control molecular aggregation. Self-assembly of three-dimensional networks with large chambers. *J. Am. Chem. Soc.* **113**, 4696–4698 (1991).
- Wuest, J. D. Engineering crystals by the strategy of molecular tectonics. *Chem. Commun.* 5830–5837 (2005).
- Hof, F., Craig, S. L., Nuckolls, C. & Rebek, J. Molecular encapsulation. *Angew. Chem. Int. Ed.* **41**, 1488–1508 (2002).
- McKeown, N. B. Nanoporous molecular crystals. *J. Mater. Chem.* **20**, 10588–10597 (2010).
- Wheeler, K. A., Grove, R. C., Davis, R. E. & Kassel, W. S. Quasiracemic materials – rediscovering Pasteur's quasiracemates. *Angew. Chem. Int. Ed.* **47**, 78–81 (2008).
- Dance, I. & Scudder, M. Supramolecular motifs: sextuple aryl embraces in crystalline  $M(2,2'-bipy)_3$  and related complexes. *Dalton Trans.* 1341–1350 (1998).
- Day, G. M., Motherwell, W. D. S. & Jones, W. Beyond the isotropic atom model in crystal structure prediction of rigid molecules: atomic multipoles versus point charges. *Cryst. Growth Des.* **5**, 1023–1033 (2005).
- Price, S. L. *et al.* Modelling organic crystal structures using distributed multipole and polarizability-based model intermolecular potentials. *Phys. Chem. Chem. Phys.* **12**, 8478–8490 (2010).
- Karamertzanis, P. G. & Pantelides, C. C. *Ab initio* crystal structure prediction. II. Flexible molecules. *Mol. Phys.* **105**, 273–291 (2007).
- Görbitz, C. H., Dalhus, B. & Day, G. M. Pseudoracemic amino acid complexes: blind predictions for flexible two-component crystals. *Phys. Chem. Chem. Phys.* **12**, 8466–8477 (2010).
- VandeVondele, J. *et al.* Quickstep: fast and accurate density functional calculations using a mixed Gaussian and plane waves approach. *Comp. Phys. Comm.* **167**, 103–108 (2005).
- Grimme, S., Antony, J., Ehrlich, S. & Krieg, H. A consistent and accurate *ab initio* parametrization of density functional dispersion correction (DFT-D) for the 94 elements H–Pu. *J. Chem. Phys.* **132**, 154104 (2010).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank EPSRC (EP/H000925/1) and the Dutch Polymer Institute for funding. The UK national high-performance computing service HECToR was used for this work through the HPC Materials Chemistry Consortium (EPSRC grant EP/F067496. A.I.C. and G.M.D. are Royal Society Wolfson Merit Award holder and A.T. and G.M.D. are Royal Society University Research Fellows.

**Author Contributions** J.T.A.J. prepared cage modules and co-crystals thereof, contributed to powder X-ray diffraction analysis, carried out high-throughput co-crystallization screening experiments and generally coordinated the experimental work; T.H. prepared cage modules and co-crystals and the nanocrystals of (**3-S**, **3-R**), and carried out microscopy; J.T.A.J. and T.H. did the gas sorption analysis; X.W. synthesized module **5**; D.J.A. contributed to the cage synthesis; J.B., M.S., S.Y.C. and A.S. performed the crystallography; K.E.J., A.T., F.C. and B.S. performed molecular and periodic simulations, in particular the DFT studies; G.M.D. led the crystal structure prediction; and A.I.C. designed the project and wrote the paper with contributions from all co-authors.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at [www.nature.com/nature](http://www.nature.com/nature). Correspondence and requests for materials should be addressed to A.I.C. ([alicoooper@liv.ac.uk](mailto:alicoooper@liv.ac.uk)).

# Control of visual cortical signals by prefrontal dopamine

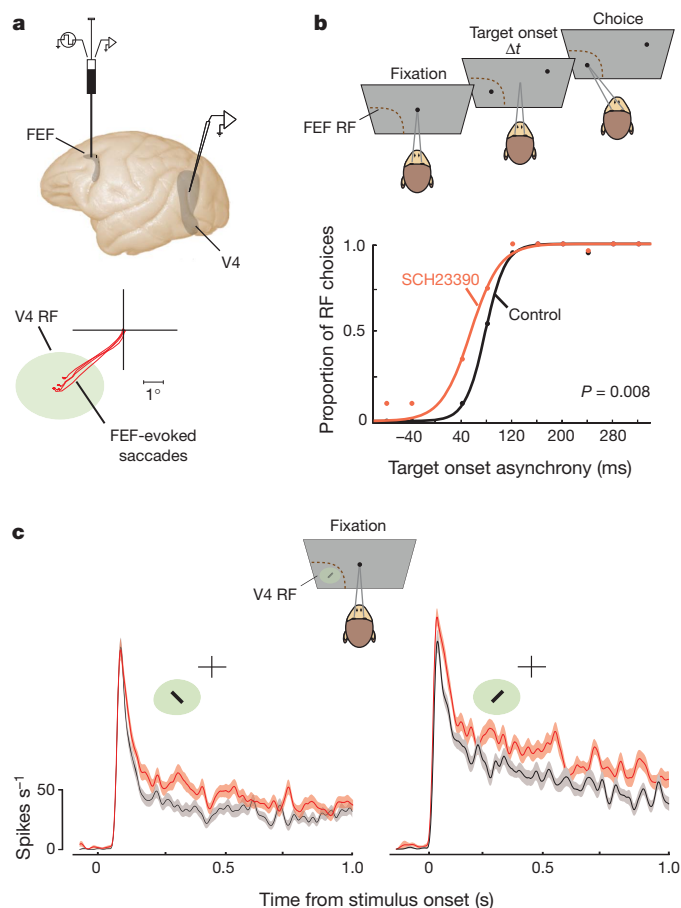
Behrad Noudoost<sup>1</sup> & Tirin Moore<sup>1</sup>

The prefrontal cortex is thought to modulate sensory signals in posterior cortices during top-down attention<sup>1,2</sup>, but little is known about the underlying neural circuitry. Experimental and clinical evidence indicate that prefrontal dopamine has an important role in cognitive functions<sup>3</sup>, acting predominantly through D1 receptors. Here we show that dopamine D1 receptors mediate prefrontal control of signals in the visual cortex of macaques (*Macaca mulatta*). We pharmacologically altered D1-receptor-mediated activity in the frontal eye field of the prefrontal cortex and measured the effect on the responses of neurons in area V4 of the visual cortex. This manipulation was sufficient to enhance the magnitude, the orientation selectivity and the reliability of V4 visual responses to an extent comparable with the known effects of top-down attention. The enhancement of V4 signals was restricted to neurons with response fields overlapping the part of visual space affected by the D1 receptor manipulation. Altering either D1- or D2-receptor-mediated frontal eye field activity increased saccadic target selection but the D2 receptor manipulation did not enhance V4 signals. Our results identify a role for D1 receptors in mediating the control of visual cortical signals by the prefrontal cortex and suggest how processing in sensory areas could be altered in mental disorders involving prefrontal dopamine.

Dopamine D1 receptors (D1Rs) are expressed by about one-quarter of all neurons in the prefrontal cortex and are localized primarily in superficial and deep layers<sup>4–6</sup>. Microiontophoretic application of the selective D1R antagonist SCH23390<sup>7</sup> at certain doses can increase the persistent, working-memory-related component of single-neuron activity in the dorsolateral prefrontal cortex<sup>3,8,9</sup>. Given the role of the prefrontal cortex in visual attention<sup>1,2</sup>, we hypothesized that D1Rs might also mediate the top-down control of visual cortical signals by the prefrontal cortex. If so, then changes in D1R-mediated prefrontal cortex activity might be sufficient to modulate signals in the posterior visual cortex, similar to the modulation observed during selective attention<sup>10</sup>. The prefrontal cortex's influence on the visual cortex is achieved in part by the frontal eye field (FEF)<sup>11,12</sup>, an oculomotor area within the posterior prefrontal cortex. The FEF has a well-established role in saccadic target selection<sup>13</sup>, but recent evidence also implicates this area in the control of spatial attention<sup>2,14,15</sup>. To test our hypothesis, we locally infused<sup>16</sup> small volumes (0.5–1  $\mu$ l) of SCH23390 into sites in the FEF of macaques performing fixation and eye movement tasks (Fig. 1a, b and Supplementary Fig. 1). We measured the effects of the FEF infusion on target selection using a free-choice saccade task<sup>17</sup>. In this task, monkeys were rewarded for choosing between two saccadic targets, one located within the FEF response field and one in the opposite hemifield. In the same experiment, we recorded the visual responses of single neurons in area V4 during fixation. In particular, we recorded neurons with response fields that overlapped the FEF response field. Thus, we tested the effects of the D1R manipulation on both visual cortical signals and saccadic target selection.

We found that altering D1R-mediated activity at FEF sites increased the tendency of monkeys to choose targets appearing within the FEF response field (Fig. 1b). In the free-choice task, the temporal onset of

the two targets was systematically varied such that the FEF response field stimulus could appear earlier or later than the opposite stimulus. A monkey's tendency to select the FEF response field target could then



**Figure 1 | Local manipulation of D1R-mediated activity in the FEF during single-neuron electrophysiology in area V4.** **a**, Lateral view of the macaque brain depicting the location of a recording microsyringe in the FEF and of recording sites in area V4. Bottom diagram shows saccades evoked by electrical microstimulation at the infusion site (red traces) and the response field (RF, green ellipse) of a recorded V4 neuron in an example experiment. **b**, Double-target saccade task used to measure the monkey's tendency to make saccades to a target within the FEF response field versus one at an opposite location across varying temporal onset asynchronies. Positive asynchrony values denote earlier onset of FEF response field targets. Bottom plot shows the leftward shift in the PES, indicating more FEF response field choices, after infusion of SCH23390 into an FEF site. **c**, Visual responses of a V4 neuron with a response field that overlapped the FEF response field, measured during passive fixation. The plot shows mean  $\pm$  s.e.m. of visual responses to a bar stimulus presented at orthogonal orientations before (grey) and after (red) the infusion of SCH23390 at the FEF site.

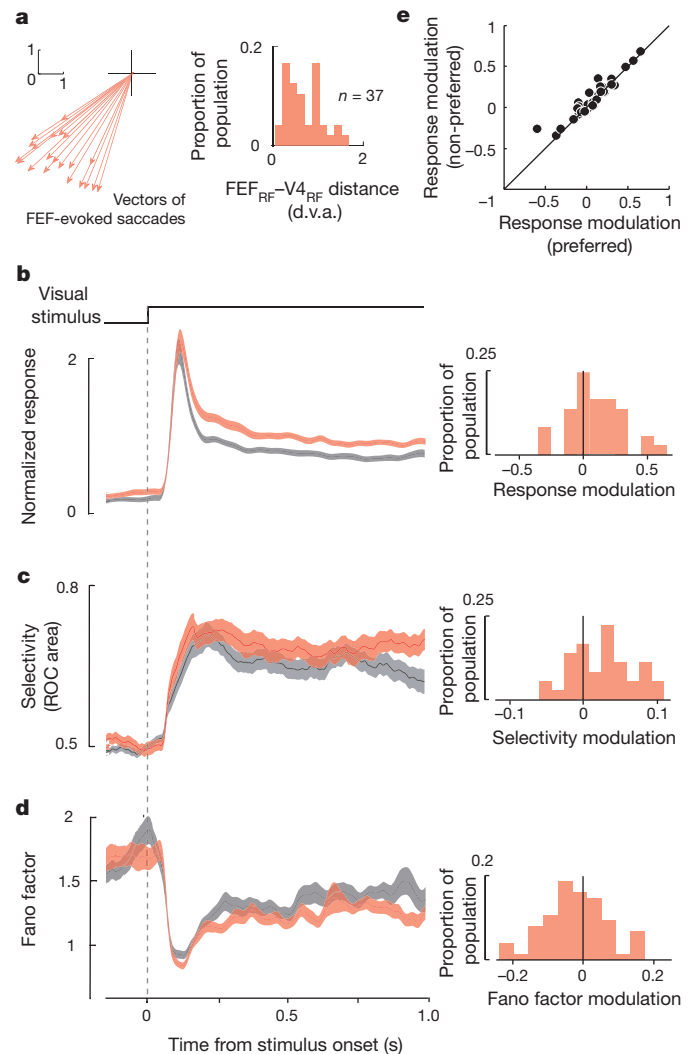
<sup>1</sup>Howard Hughes Medical Institute and Department of Neurobiology, Stanford University School of Medicine, Fairchild building, 299 Campus Drive West, Stanford, California 94305, USA.

be measured as the temporal onset asynchrony required for an equal probability of selecting either stimulus; we termed this the point of equal selection (PES). In the example experiment shown, the monkey chose the FEF response field target as often as the opposite target when the former appeared 76 ms earlier (PES = 76). However, infusion of SCH23390 (0.85  $\mu$ l) into the FEF reduced the PES by 23 ms (binary logistic regression,  $P = 0.007$ ), thereby increasing the proportion of FEF response field target choices.

In the same experiment, we also measured the responses of V4 neurons to oriented bars during fixation in a separate task (Fig. 1c and Supplementary Methods). We found that the increase in target selection after the SCH23390 infusion was accompanied by an enhanced V4 neuronal response to oriented bars appearing within the overlapping V4 and FEF response fields. The example neuron shown was selective for orientation: it responded more to the 45° than to the 135° bar stimulus ( $P < 10^{-3}$ ). After the infusion of SCH23390, there was a significant increase in the overall visual response of this neuron as well as a significant increase in the differential response to the two orientations (two-way analysis of variance, SCH23390 effect,  $P < 10^{-3}$ ; SCH23390–orientation interaction,  $P < 10^{-3}$ ). Thus, the local perturbation of D1R-mediated FEF activity not only caused the monkey to select FEF response field stimuli as saccade targets more frequently, it also led to enhanced and more selective visual responses of a V4 neuron representing the same part of space.

We studied the visual responses of 37 V4 neurons with response fields that overlapped the response fields of FEF infusion sites. The average (mean  $\pm$  s.e.m.) distance between V4 response field and FEF response field centres was  $0.71 \pm 0.07$  degrees of visual angle (d.v.a.) (Fig. 2a). As with the example neuron, we measured the responses of all neurons to oriented bars appearing in their response field during a 1 s fixation period (Fig. 2b). Before the onset of the visual stimulus, there was a significant elevation in baseline activity after the D1R manipulation ( $\Delta$  baseline =  $0.077 \pm 0.186$ ,  $P = 0.030$ ). In addition to the baseline increase, the visually driven response of V4 neurons was enhanced by 17% above the control response ( $\Delta$  response =  $0.121 \pm 0.054$ ,  $P = 0.018$ ). We confirmed that the enhancement in the visual response was not due to systematic changes in eye position during stimulus presentation (Supplementary Fig. 2). The enhancement of the visual response was independently significant for both preferred ( $\Delta$  preferred =  $0.264 \pm 0.087$ ;  $P = 0.004$ ) and non-preferred stimuli ( $\Delta$  non-preferred =  $0.132 \pm 0.062$ ;  $P = 0.032$ ). There was also an increase in the response difference between the preferred and non-preferred orientations ( $\Delta$  response difference =  $0.132 \pm 0.041$ ;  $P = 0.004$ ) (Supplementary Fig. 3), indicating an increase in orientation selectivity. To measure selectivity more quantitatively, we used a receiver-operating characteristic (ROC) analysis to quantify the degree to which each neuron's responses could be used to judge stimulus orientation (Fig. 2c). This analysis confirmed that V4 neurons were more orientation selective after changes in D1R-mediated FEF activity ( $\Delta$  ROC area =  $0.035 \pm 0.009$ ,  $P < 10^{-3}$ ). The enhancement in the magnitude and selectivity of the V4 response was accompanied by a decrease in the trial-to-trial variability of visual responses. We measured the variability of V4 responses across trials by computing the Fano factor, which is the variance in the spike count divided by its mean. We found that the Fano factor of V4 responses was reduced after the D1R manipulation ( $\Delta$  FF =  $-0.105 \pm 0.045$ ;  $P < 10^{-3}$ ) (Fig. 2d and Supplementary Fig. 4). All three V4 effects were comparable in magnitude to the known effects of top-down attention and consistent with a multiplicative increase in the gain of visual signals<sup>18,19</sup> (Fig. 2e).

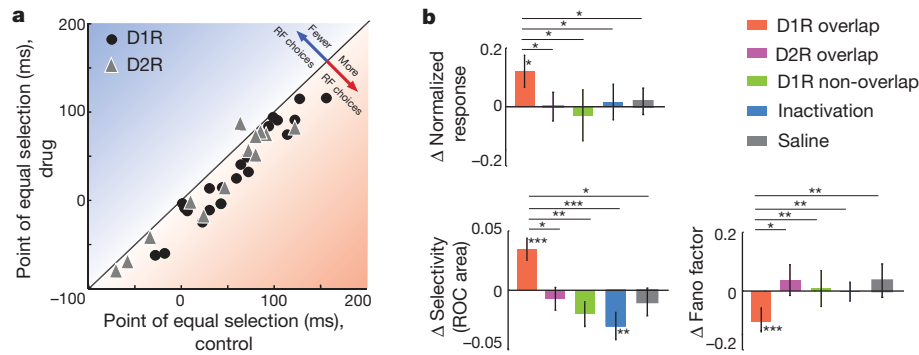
The effect of the D1R manipulation on saccadic target selection was highly consistent across the two monkeys tested. In 21 double-target experiments, the PES was reduced in every case (Fig. 3a). The mean PES shifted in favour of the FEF response field stimulus by an average of 27 ms ( $\Delta$  PES =  $-26.934 \pm 3.086$ ,  $P < 10^{-3}$ ), significantly increasing the overall proportion of FEF response field choices



**Figure 2 | Manipulation of D1R-mediated activity enhances V4 visual signals.** **a**, Average vectors of saccades evoked at all FEF sites that overlapped V4 response fields (left panel). The distribution of distances between the endpoints of evoked saccades and the centres of overlapping V4 response fields for 37 V4 neurons is shown in the right panel. **b–d**, The mean normalized response magnitude (**b**), orientation selectivity (**c**) and response variability (Fano factor) (**d**) of V4 neurons before (grey) and after (red) microinfusion of SCH23390 into the FEF. Means  $\pm$  s.e.m. are shown within a 100-ms moving window measured during the 1-s response field stimulus presentation (top event plot). Histograms to the right of each response profile show the distributions of modulation indices for response magnitude (**b**), selectivity (**c**) and variability (**d**) across the population of neurons. **e**, Comparison of V4 response modulation after the SCH23390 infusion for preferred and non-preferred response field stimuli.

(chi-squared = 80.60,  $P < 10^{-3}$ ) and thus indicating that the D1R manipulation increased the monkeys' tendency to target FEF response field stimuli. The increase in target selection was apparent across a range of drug dosages (Supplementary Fig. 5). In addition to the D1R manipulation, we tested the effects of the D2R agonist quinpirole. Previous studies using this drug found that it does not affect persistent activity but rather increases saccade-related activity within the dorsolateral prefrontal cortex<sup>20</sup>. We found that local manipulation of D2R-mediated FEF activity, like the D1R manipulation, increased the selection of FEF response field targets (Fig. 3a). The PES shifted by an average of 22 ms ( $\Delta$  PES =  $-21.993 \pm 6.758$ ,  $P = 0.010$ ), increasing the proportion of FEF response field choices (chi-squared = 13.86,  $P < 10^{-3}$ ). Thus, the D1R- and D2R-mediated manipulations of FEF activity resulted in equivalent increases in saccadic target selection.





**Figure 3 | Changes in saccadic target selection and V4 visual responses.** **a**, Scatter plot shows the consistent increase in FEF response field target choices (decrease in PES) after manipulation of both D1R-mediated (circles) and D2R-mediated (triangles) FEF activity. For both drug effects, the increase in FEF response field target selection was constant across a range of control PES values; the slope in the linear fit did not differ significantly from unity in either case

Despite the increase in target selection, manipulation of D2R-mediated activity in the FEF failed to enhance the responses of V4 neurons. We found no significant effect on the visual response magnitude, orientation selectivity or response variability of V4 neurons after the D2R manipulation ( $\Delta \text{response} = 0.001 \pm 0.048$ ,  $P = 0.999$ ;  $\Delta \text{ROC area} = -0.007 \pm 0.010$ ,  $P = 0.426$ ;  $\Delta \text{FF} = 0.037 \pm 0.052$ ,  $P = 0.338$ ;  $n = 15$ ) (Fig. 3b). Moreover, the changes in these measures were all significantly different from the changes we observed after the D1R manipulation ( $\Delta \text{response}_{\text{D2R}} < \Delta \text{response}_{\text{D1R}}$ ,  $P = 0.045$ ;  $\Delta \text{selectivity}_{\text{D2R}} < \Delta \text{selectivity}_{\text{D1R}}$ ,  $P = 0.011$ ;  $\Delta \text{FF}_{\text{D2R}} > \Delta \text{FF}_{\text{D1R}}$ ,  $P = 0.019$ ). Thus, the equivalent effects of D1R and D2R manipulations on saccadic target selection were accompanied by contrasting effects in V4, with the enhancement of visual signals being specific to D1R-mediated activity. We also found that this enhancement was confined to V4 neurons with response fields that overlapped the FEF response field. For V4 neurons with response fields that did not overlap the FEF response field (mean distance between V4 response field and FEF response field =  $9.00 \pm 0.86$  d.v.a.;  $n = 15$ ), we found no significant effect of the D1R manipulation on response magnitude ( $\Delta \text{response} = -0.028 \pm 0.087$ ,  $P = 0.9780$ ), orientation selectivity ( $\Delta \text{ROC area} = -0.017 \pm 0.010$ ,  $P = 0.187$ ) or the Fano factor ( $\Delta \text{FF} = 0.010 \pm 0.043$ ,  $P = 0.688$ ). Of note, the changes in these measures were all significantly different from the changes observed in neurons with overlapping response fields ( $\Delta \text{response}_{\text{non-overlap}} < \Delta \text{response}_{\text{overlap}}$ ,  $P = 0.044$ ;  $\Delta \text{selectivity}_{\text{non-overlap}} < \Delta \text{selectivity}_{\text{overlap}}$ ,  $P = 0.007$ ;  $\Delta \text{FF}_{\text{non-overlap}} > \Delta \text{FF}_{\text{overlap}}$ ,  $P = 0.034$ ) (Fig. 3b). Thus, the enhancement in visual cortical signalling produced by manipulation of D1R-mediated FEF activity was spatially specific.

We also tested the effect of complete inactivation of FEF sites on the responses of V4 neurons with overlapping response fields. Previous studies have shown that local inactivation of the FEF disrupts saccadic target selection and impairs attention<sup>17,21</sup>. We therefore wondered if inactivation could reduce the components of V4 responses that were enhanced by the D1R manipulation. We locally inactivated FEF sites using the GABA<sub>A</sub> ( $\gamma$ -aminobutyric acid subtype A) receptor agonist muscimol. Unlike the sparsely expressed D1Rs, GABA<sub>A</sub> receptors are expressed by all neurons in all cortical layers<sup>22</sup>. As in previous studies, local inactivation of FEF sites with muscimol decreased the targeting of FEF response field stimuli. It also significantly reduced V4 orientation selectivity ( $\Delta \text{ROC area} = -0.030 \pm 0.011$ ,  $P = 0.003$ ;  $n = 33$ ). However, the inactivation did not change the response magnitude or variability of V4 neurons ( $\Delta \text{response} = 0.016 \pm 0.061$ ,  $P = 0.809$ ;  $\Delta \text{FF} = -0.002 \pm 0.023$ ,  $P = 0.921$ ) (Fig. 3b). Thus, in contrast to the D1R manipulation which altered all three components of V4 activity, complete inactivation altered only one. All three inactivation effects were significantly different from the D1R effects

(D1R: slope = 0.96,  $P = 0.552$ ; D2R: slope = 0.97,  $P = 0.502$ ). **b**, Changes in response magnitude, orientation selectivity and response variability (Fano factor) after each drug manipulation. Changes shown are mean differences from pre-infusion values. Error bars denote s.e.m.; \*,  $P < 0.05$ ; \*\*,  $P < 0.01$ ; \*\*\*,  $P < 0.001$ .

( $\Delta \text{response}_{\text{muscimol}} < \Delta \text{response}_{\text{D1R}}$ ,  $P = 0.024$ ;  $\Delta \text{selectivity}_{\text{muscimol}} < \Delta \text{selectivity}_{\text{D1R}}$ ,  $P < 10^{-3}$ ;  $\Delta \text{FF}_{\text{muscimol}} > \Delta \text{FF}_{\text{D1R}}$ ,  $P = 0.007$ ). Although the reduction in orientation selectivity is consistent with previous electrical microstimulation studies<sup>12</sup> and with the effects of inactivation on orientation discrimination<sup>21</sup>, the lack of a reduction in response magnitude may seem inconsistent. However, we suggest that this difference is due to variation between experimental paradigms (Supplementary Discussion). Finally, we tested for any effect of vehicle (saline) infusion into the FEF. The infusion of saline failed to change the response magnitude, selectivity or variability of V4 neurons ( $\Delta \text{response} = 0.018 \pm 0.048$ ,  $P = 0.380$ ;  $\Delta \text{ROC area} = -0.010 \pm 0.013$ ,  $P = 0.569$ ;  $\Delta \text{FF} = -0.035 \pm 0.061$ ,  $P = 0.179$ ;  $n = 12$ ) (Fig. 3b). All three measures were significantly different from the D1R effects ( $\Delta \text{response}_{\text{saline}} < \Delta \text{response}_{\text{D1R}}$ ,  $P = 0.045$ ;  $\Delta \text{selectivity}_{\text{saline}} < \Delta \text{selectivity}_{\text{D1R}}$ ,  $P = 0.013$ ;  $\Delta \text{FF}_{\text{saline}} > \Delta \text{FF}_{\text{D1R}}$ ,  $P = 0.009$ ).

Our results identify prefrontal D1Rs as a component of the neural circuitry controlling signals in the visual cortex. Manipulation of D1R-mediated FEF activity was sufficient to enhance the magnitude, reliability and visual selectivity of neuronal responses in area V4, three known effects of visual attention. The observed enhancement might account for the benefits in visually guided behaviour that accompany attentional deployment (Supplementary Fig. 6), although a causal link between attentional modulation of visual cortical signals and visual perception remains to be established. We have demonstrated that visual representations in posterior areas can be altered merely by changes in dopamine tone in the prefrontal cortex. Given the complex effects of dopamine through D1Rs, one might predict that at 'optimum' dopamine levels<sup>9</sup>, optimal top-down control of visual cortical signals would be achieved.

The circuitry underlying top-down control of the visual cortex probably involves several different neuromodulators<sup>23</sup> and an array of different brain structures<sup>24</sup>. Our results show that this circuitry involves prefrontal dopamine acting via D1Rs. In the dorsolateral prefrontal cortex, dopamine D1Rs are thought to modulate recurrent glutamatergic connections, thereby influencing activity related to working memory in this area<sup>25,26</sup>. This study shows that D1Rs contribute to the FEF's control of visual signals by an analogous mechanism, namely by modulating long-range, recurrent connections between the FEF and the visual cortex (Supplementary Fig. 7). Because FEF neurons in the superficial layer are reciprocally connected with neurons in V4<sup>2,27</sup>, dopaminergic modulation of these connections via D1Rs in the superficial layer would be expected to mediate the FEF's control of V4 signals. The specificity of V4 effects to D1Rs, rather than D2Rs, might be explained by the relative absence of D2Rs in superficial layers of the prefrontal cortex<sup>4-6</sup>. The equivalent effects of D1R and D2R

manipulations on target selection might be explained by the presence of both receptor subtypes in infragranular layers of the cortex<sup>4–6</sup>, where layer-V FEF neurons project to the superior colliculus<sup>27</sup>.

Impairments in saccadic control are prominent among the impairments exhibited in attention deficit/hyperactivity disorder (ADHD)<sup>28</sup>. The observed influence of prefrontal DIRs on saccadic target selection and visual cortical signals, combined with their known influence on persistent activity, may explain the behavioural links between saccadic control, attention and working memory<sup>29</sup> and the coincidence of their corresponding impairments in ADHD<sup>30</sup>.

## METHODS SUMMARY

The effects of pharmacological perturbation of FEF activity on target selection and the visual responses of V4 neurons were studied in three macaques (*Macaca mulatta*) performing fixation and eye movement tasks (Supplementary Methods). All experimental procedures were in accordance with the National Institutes of Health guide for the care and use of laboratory animals and with the Society for Neuroscience guidelines and policies. They were also approved by the Stanford University animal care and use committee. Eye position was monitored with a scleral search coil. In each experiment, we infused small volumes of drug into sites in the FEF through a surgically implanted titanium chamber overlying the arcuate sulcus using a custom-made recording microinjector. We identified FEF sites by eliciting short-latency, fixed-vector saccadic eye movements with trains (50–100 ms) of biphasic current pulses ( $\leq 50 \mu\text{A}$ ; 250 Hz; 0.25 ms duration). In the same experiment, recordings from V4 neurons were made through a chamber overlying the prelunate gyrus. Response fields of V4 neurons were all located in the lower quadrant of the contralateral hemifield ( $<12^\circ$  eccentricity). The position of the FEF microinjector was adjusted so that the saccade elicited by FEF microstimulation shifted the monkey's gaze either to within the V4 response field (overlapping) or far outside it (non-overlapping).

Received 20 May 2010; accepted 14 March 2011.

Published online 15 May 2011.

- Miller, E. K. & Cohen, J. D. An integrative theory of prefrontal cortex function. *Annu. Rev. Neurosci.* **24**, 167–202 (2001).
- Noudoost, B., Chang, M. H., Steinmetz, N. A. & Moore, T. Top-down control of visual attention. *Curr. Opin. Neurobiol.* **20**, 183–190 (2010).
- Robbins, T. W. & Arnsten, A. F. The neuropsychopharmacology of fronto-executive function: monoaminergic modulation. *Annu. Rev. Neurosci.* **32**, 267–287 (2009).
- Lidow, M. S., Goldman-Rakic, P. S., Gallager, D. W. & Rakic, P. Distribution of dopaminergic receptors in the primate cerebral cortex: quantitative autoradiographic analysis using [<sup>3</sup>H]raclopride, [<sup>3</sup>H]spiperone and [<sup>3</sup>H]SCH23390. *Neuroscience* **40**, 657–671 (1991).
- Lidow, M. S., Wang, F., Cao, Y. & Goldman-Rakic, P. S. Layer V neurons bear the majority of mRNAs encoding the five distinct dopamine receptor subtypes in the primate prefrontal cortex. *Synapse* **28**, 10–20 (1998).
- Santana, N., Mengod, G. & Artigas, F. Quantitative analysis of the expression of dopamine D1 and D2 receptors in pyramidal and GABAergic neurons of the rat prefrontal cortex. *Cereb. Cortex* **19**, 849–860 (2009).
- Bourne, J. A. SCH 23390: the first selective dopamine D1-like receptor antagonist. *CNS Drug Rev.* **7**, 399–414 (2001).
- Williams, G. V. & Goldman-Rakic, P. S. Modulation of memory fields by dopamine D1 receptors in prefrontal cortex. *Nature* **376**, 572–575 (1995).
- Vijayraghavan, S., Wang, M., Birnbaum, S. G., Williams, G. V. & Arnsten, A. F. Inverted-U dopamine D1 receptor actions on prefrontal neurons engaged in working memory. *Nature Neurosci.* **10**, 376–384 (2007).
- Reynolds, J. H. & Chelazzi, L. Attentional modulation of visual processing. *Annu. Rev. Neurosci.* **27**, 611–647 (2004).
- Gregoriou, G. G., Gotts, S. J., Zhou, H. & Desimone, R. High-frequency, long-range coupling between prefrontal and visual cortex during attention. *Science* **324**, 1207–1210 (2009).
- Moore, T. & Armstrong, K. M. Selective gating of visual signals by microstimulation of frontal cortex. *Nature* **421**, 370–373 (2003).
- Schall, J. D. & Hanes, D. P. Neural basis of saccade target selection in frontal eye field during visual search. *Nature* **366**, 467–469 (1993).
- Thompson, K. G., Biscoe, K. L. & Sato, T. R. Neuronal basis of covert spatial attention in the frontal eye field. *J. Neurosci.* **25**, 9479–9487 (2005).
- Armstrong, K. M., Chang, M. H. & Moore, T. Selection and maintenance of spatial information by frontal eye field neurons. *J. Neurosci.* **29**, 15621–15629 (2009).
- Noudoost, B. & Moore, T. A reliable microinjector system for use in behaving monkeys. *J. Neurosci. Methods* **194**, 218–223 (2011).
- Schiller, P. H. & Tehovnik, E. J. Cortical inhibitory circuits in eye-movement generation. *Eur. J. Neurosci.* **18**, 3127–3133 (2003).
- Mitchell, J. F., Sundberg, K. A. & Reynolds, J. H. Differential attention-dependent response modulation across cell classes in macaque visual area V4. *Neuron* **55**, 131–141 (2007).
- McAdams, C. J. & Maunsell, J. H. Effects of attention on orientation-tuning functions of single neurons in macaque cortical area V4. *J. Neurosci.* **19**, 431–441 (1999).
- Wang, M., Vijayraghavan, S. & Goldman-Rakic, P. S. Selective D2 receptor actions on the functional circuitry of working memory. *Science* **303**, 853–856 (2004).
- Monosov, I. E. & Thompson, K. G. Frontal eye field activity enhances object identification during covert visual search. *J. Neurophysiol.* **102**, 3656–3672 (2009).
- Huntsman, M. M., Isackson, P. J. & Jones, E. G. Lamina-specific expression and activity-dependent regulation of seven GABA<sub>A</sub> receptor subunit mRNAs in monkey visual cortex. *J. Neurosci.* **14**, 2236–2259 (1994).
- Herrero, J. L. et al. Acetylcholine contributes through muscarinic receptors to attentional modulation in V1. *Nature* **454**, 1110–1114 (2008).
- McAlonan, K., Cavanaugh, J. & Wurtz, R. H. Guarding the gateway to cortex with attention in visual thalamus. *Nature* **456**, 391–394 (2008).
- Goldman-Rakic, P. S. Cellular basis of working memory. *Neuron* **14**, 477–485 (1995).
- Durstewitz, D., Seamans, J. K. & Sejnowski, T. J. Neurocomputational models of working memory. *Nature Neurosci.* **3** (Suppl.), 1184–1191 (2000).
- Pouget, P. et al. Visual and motor connectivity and the distribution of calcium-binding proteins in macaque frontal eye field: implications for saccade target selection. *Front. Neuroanat.* **3**, 1–14 (2009).
- Munoz, D. P., Armstrong, I. T., Hampton, K. A. & Moore, K. D. Altered control of visual fixation and saccadic eye movements in attention-deficit hyperactivity disorder. *J. Neurophysiol.* **90**, 503–514 (2003).
- Awh, E., Armstrong, K. M. & Moore, T. Visual and oculomotor selection: links, causes and implications for spatial attention. *Trends Cogn. Sci.* **10**, 124–130 (2006).
- Castellanos, F. X. & Tannock, R. Neuroscience of attention-deficit/hyperactivity disorder: the search for endophenotypes. *Nature Rev. Neurosci.* **3**, 617–628 (2002).

Supplementary Information is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank D. S. Aldrich for technical assistance, N. Steinmetz for help with Fano factor analysis and W. T. Newsome, E. I. Knudsen, K. M. Armstrong and R. F. Squire for comments on the manuscript. This work was supported by NIH EY014924, NSF IOB-0546891, The McKnight Foundation and an IBRO Fellowship to B.N.

**Author Contributions** B.N. designed and performed experiments, analysed data and wrote the paper; T.M. designed and performed experiments and wrote the paper.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at [www.nature.com/nature](http://www.nature.com/nature). Correspondence and requests for materials should be addressed to B.N. ([behrad@stanford.edu](mailto:behrad@stanford.edu)).

# Forces between clustered stereocilia minimize friction in the ear on a subnanometre scale

Andrei S. Kozlov<sup>1</sup>, Johannes Baumgart<sup>2</sup>, Thomas Risler<sup>3,4,5</sup>, Corstiaan P. C. Versteegh<sup>1,6</sup> & A. J. Hudspeth<sup>1</sup>

The detection of sound begins when energy derived from an acoustic stimulus deflects the hair bundles on top of hair cells<sup>1</sup>. As hair bundles move, the viscous friction between stereocilia and the surrounding liquid poses a fundamental physical challenge to the ear's high sensitivity and sharp frequency selectivity. Part of the solution to this problem lies in the active process that uses energy for frequency-selective sound amplification<sup>2,3</sup>. Here we demonstrate that a complementary part of the solution involves the fluid–structure interaction between the liquid within the hair bundle and the stereocilia. Using force measurement on a dynamically scaled model, finite-element analysis, analytical estimation of hydrodynamic forces, stochastic simulation and high-resolution interferometric measurement of hair bundles, we characterize the origin and magnitude of the forces between individual stereocilia during small hair-bundle deflections. We find that the close apposition of stereocilia effectively immobilizes the liquid between them, which reduces the drag and suppresses the relative squeezing but not the sliding mode of stereociliary motion. The obliquely oriented tip links couple the mechanotransduction channels to this least dissipative coherent mode, whereas the elastic horizontal top connectors that stabilize the structure further reduce the drag. As measured from the distortion products associated with channel gating at physiological stimulation amplitudes of tens of nanometres, the balance of viscous and elastic forces in a hair bundle permits a relative mode of motion between adjacent stereocilia that encompasses only a fraction of a nanometre. A combination of high-resolution experiments and detailed numerical modelling of fluid–structure interactions reveals the physical principles behind the basic structural features of hair bundles and shows quantitatively how these organelles are adapted to the needs of sensitive mechanotransduction.

A hair bundle is a microscopic array of quasi-rigid, cylindrical stereocilia separated by small gaps filled with viscous endolymph. Like an array of organ pipes, the stereocilia vary monotonically in length across the hair bundle (Supplementary Information section 1). The tip of each short stereocilium is attached to the side of the longest adjacent stereocilium by a tip link, the tension in which controls the opening and closing of transduction channels. Adjacent stereocilia are also interconnected along all three hexagonal axes by horizontal top connectors. At the tall edge of the bundle in many species stands a single kinocilium, the process to which mechanical stimuli are applied and that is ligated to the adjacent stereocilia by kinociliary links.

When a solid object such as a hair bundle moves through a viscous fluid, the interplay between viscosity and inertia produces a spatial gradient of fluid velocity and the shear between successive layers of fluid causes friction<sup>4</sup>. The characteristic decay length of the shear waves created by an oscillating body scales as  $\sqrt{\eta/(\omega\rho)}$ , in which  $\eta$  is the fluid's dynamic viscosity,  $\rho$  is its density and  $\omega$  is the angular frequency of motion<sup>5</sup>. Because this length scale greatly exceeds the

distance between stereocilia, viscous forces can couple all motions within a hair bundle. On the other hand, the pivotal stiffness of individual stereociliary rootlets opposes deflection. The viscous forces in the endolymph, elastic forces in the stereociliary pivots and links, and (at high frequencies) inertial forces associated with the liquid and stereociliary masses together determine all the motions within a bundle.

Although stereociliary motion can be measured directly with an interferometer (Supplementary Information section 1), a qualitative appreciation of the liquid's movement can be obtained from the associated drag. When a fluid moves between nearby cylinders with axes perpendicular to the flow, the drag on each cylinder exceeds that on an identical cylinder placed alone in a flow with the same average velocity. At a Reynolds number well below one, this effect is strong and long-range<sup>6,7</sup>. One might therefore expect a drag coefficient for a hair bundle several hundred times that of an isolated stereocilium. Instead, the measured values are of similar magnitude: for six interferometric measurements in each case, the drag coefficient for a single stereocilium is  $16 \pm 5 \text{ nN s m}^{-1}$ , whereas that for an entire bundle lacking tip links is only  $30 \pm 13 \text{ nN s m}^{-1}$ . Because we determined the drag coefficient for hair bundles that lacked tip links and displayed coherent Brownian motion, the latter value is about a quarter of that typically reported in the literature<sup>8</sup>. We note that these values resemble those calculated for geometrical solids of similar dimensions pivoting at their bases and evaluated at their tips<sup>9,10</sup>:  $14 \text{ nN s m}^{-1}$  for a cylinder of the size of a stereocilium and  $29 \text{ nN s m}^{-1}$  for a hemi-ellipsoid with the dimensions of a hair bundle. The small difference between the drag coefficients for a single stereocilium and for an entire hair bundle reveals the striking advantage that grouping stereocilia in a tightly packed array offers to the auditory system.

Although stereocilia may slide past each other quite easily, large forces are required to squeeze them together or separate them. To estimate these forces, we constructed a macroscopic model of a hair bundle with the surrounding liquid, preserving the scaling between the physical quantities of importance (Supplementary Information section 2). A simplified model of a bullfrog's hair bundle enlarged 12,000 times was placed in a 2.2% solution of methylcellulose, which is 5,000 times as viscous as water. A single stereocilium was pulled at speeds of  $0.015\text{--}1.11 \text{ mm s}^{-1}$  while the frictional force was measured. After rescaling the time, length and mass values to those of a biological hair bundle, we estimated the drag coefficient for the small-gap separation of a single stereocilium to be  $1,000\text{--}10,000 \text{ nN s m}^{-1}$ , which is several hundred times that for the movement of an isolated stereocilium. This order-of-magnitude demonstration confirmed that very large frictional forces oppose the squeezing motion, indicating the importance of hydrodynamics in the coupling of stereocilia.

Elastic forces become dominant in the low-frequency regime and inertial forces become dominant in the high-frequency regime of hair-bundle motion. To quantify the forces as a function of frequency, we developed a finite-element model in which we could manipulate the mechanical properties of the elastic links while explicitly representing the liquid around and between the stereocilia (Supplementary

<sup>1</sup>Howard Hughes Medical Institute and Laboratory of Sensory Neuroscience, The Rockefeller University, 1230 York Avenue, New York, New York 10065, USA. <sup>2</sup>Institute of Scientific Computing, Department of Mathematics, Technische Universität Dresden, 01062 Dresden, Germany. <sup>3</sup>Institut Curie, Centre de Recherche, F-75005 Paris, France. <sup>4</sup>UPMC Université Paris 06, UMR 168, F-75005 Paris, France. <sup>5</sup>CNRS, UMR 168, F-75005 Paris, France. <sup>6</sup>Experimental Zoology Group, Wageningen University, 6709 PG Wageningen, The Netherlands.



Information section 3). The model has about 800,000 degrees of freedom and is the first finite-element model to resolve the liquid motion in the gaps between stereocilia as well as in the outer boundary layer. The hair bundle is excited in the model by imposing an oscillatory displacement at varying frequencies on the kinocilium.

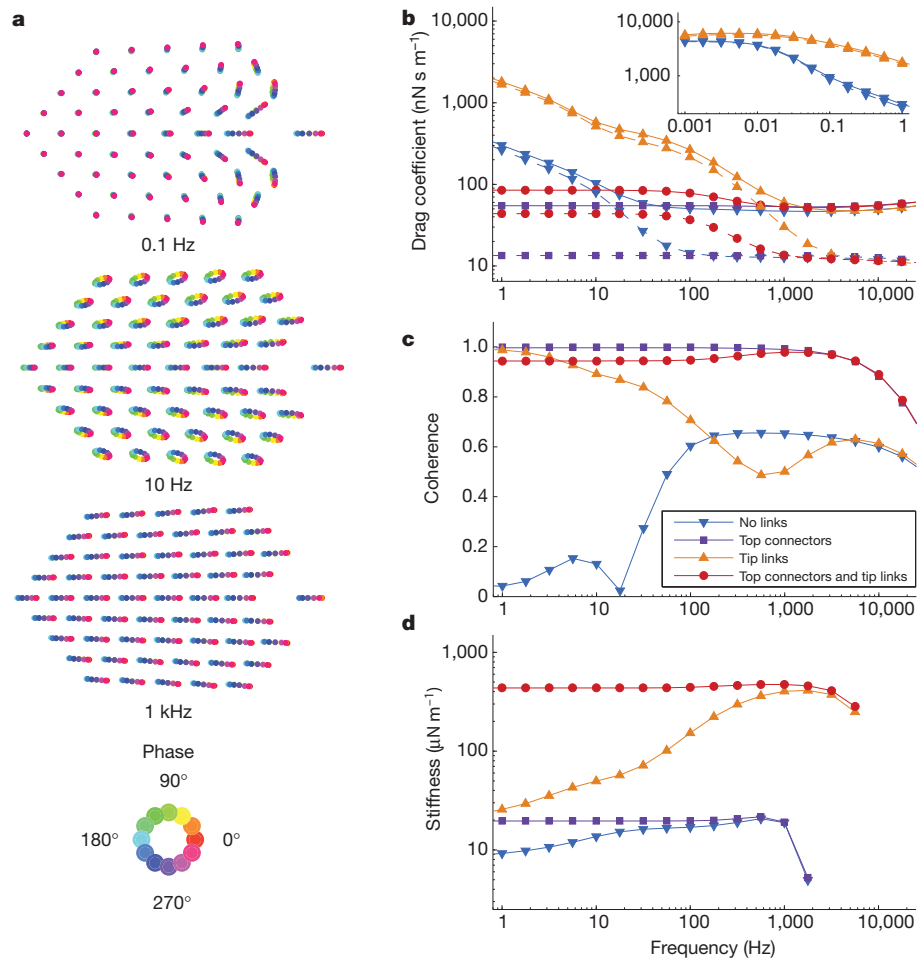
First, we examined the model including only pivotal stiffness, hydrodynamic drag and inertial mass (Supplementary Movie 1). At low frequencies, the viscous force is small and only the stimulated kinocilium and its tightly joined next neighbours move (Fig. 1a). The associated drag coefficient is about  $5,000 \text{ nN s m}^{-1}$  (Fig. 1b inset), a value in agreement with the result obtained with the scaled dynamical model. Because frictional forces increase linearly with frequency whereas elastic coupling remains constant for a given displacement, hydrodynamic coupling progressively entrains the whole hair bundle at higher frequencies (Fig. 1a). As the squeezing modes subside, the drag coefficient per stereocilium decreases, dropping by two orders of magnitude by 100 Hz (Fig. 1b). Above that frequency the entire bundle moves as a unit (Fig. 1a).

Exciting the hair bundle and recording the linear responses at its opposite edges allowed us to compute the coherence of motion, a quantity that could be directly compared with interferometric measurements<sup>11</sup> (Supplementary Information sections 1, 3 and 4). A hair bundle in the finite-element model without any interstereociliary linkages

displays a coherence exceeding 0.6 between 100 Hz and 5 kHz until inertia intervenes at higher frequencies (Fig. 1c). Adding horizontal top connectors with a stiffness of  $20 \text{ mN m}^{-1}$  to the model strongly increases the coherence, especially at low frequencies, and reduces the drag (Fig. 1b and c and Supplementary Movie 2). This value for the stiffness of top connectors was chosen such that the output coherence spectrum matched the experimental observations. It is corroborated by the distortion-product experiments discussed below and accords with published experimental and modelling studies<sup>12–14</sup>.

Adding to the model tip links with a stiffness of  $1 \text{ mN m}^{-1}$ , rather than top connectors, introduces some elastic coupling between the stereocilia of a given column (Fig. 1c and Supplementary Movie 3), but this coupling is inefficient. For low frequencies at which hydrodynamic coupling is weak, only the excited column moves significantly. Moreover, because they are oriented obliquely, the tip links pull the stereocilia towards one another during positive deflections and allow them to separate during the complementary half-cycles. Both effects dramatically increase the drag, which originates almost entirely from the liquid within the hair bundle (Fig. 1b).

Including both horizontal top connectors and tip links in the model increases the coherence for all frequencies below 5 kHz to 0.94 (Fig. 1c and Supplementary Movie 4), a value comparable to the experimental measurement. This model displays a low drag coefficient of



**Figure 1 | Finite-element analysis of fluid–structure interactions in a hair bundle.** **a**, Three top views illustrate the calculated motion of a hair bundle without elastic elements other than the kinociliary links and rootlets in response to sinusoidal deflections of the kinocilium, which lies at the right in each diagram. The colour scale (at the bottom) identifies successive positions through one cycle of stimulation with phase progressing counterclockwise. As the frequency increases, the stereocilia display a transition from weakly coupled to collective motion. The frequency dependence of the drag coefficient (**b**), the coherence (**c**) and the stiffness (**d**) are obtained from the model with four

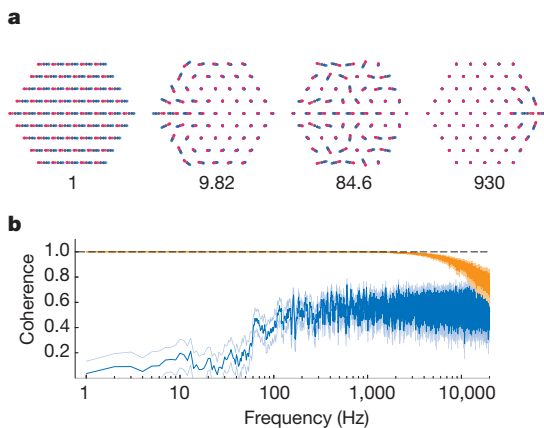
configurations of the coupling between stereocilia: with only pivotal stiffness and hydrodynamic drag (blue downtriangles); adding horizontal top connectors with a stiffness of  $20 \text{ mN m}^{-1}$  (purple squares); adding instead tip links with a stiffness of  $1 \text{ mN m}^{-1}$  (orange uptriangles); and adding both top connectors and tip links (red circles). The drag coefficient in **b** was calculated in the presence of liquid both outside and inside the hair bundle (solid lines) as well as with the liquid inside only (dashed lines). The inset in **b**, which has axis labels identical to those of the main panel, displays the behaviour of two model configurations at extremely low frequencies.

$85 \text{ nN s m}^{-1}$  that changes little with frequency (Fig. 1b), with the drag originating primarily from the external liquid but with some contribution from relative motions in the bundle, and a stiffness of  $450 \mu\text{N m}^{-1}$  (Fig. 1d), similar to that reported for intact hair bundles<sup>8,15</sup>. We note that at frequencies below 1 kHz the tip links strongly increase the hair bundle's drag, whereas the top connectors largely suppress this effect. At higher frequencies, the liquid alone provides such a strong coupling that the tip links do not affect the drag significantly. This frequency-dependent transition between elastic and viscous regimes might explain why some high-frequency hair cells, in particular mammalian inner hair cells, apparently lack top connectors<sup>16</sup>.

We next explored the fluid–structure interactions in an analytically tractable and intrinsically stochastic model that allowed us to generate time series that could be compared directly with experiments (Supplementary Information section 5). Unlike the harmonic single-point excitation in the finite-element model, the movement in this instance was caused by the coupling of each individual stereocilium to the thermal bath through a Langevin equation. The movements between each pair of stereocilia were derived from a basis set of elementary motions, for which we solved the Stefan–Reynolds equations within the lubrication approximation (Supplementary Information section 6).

Setting the elastic coupling to zero, we obtained a damping matrix with eigenvalues spanning about three orders of magnitude from the least damped collective modes to the most damped relative ones (Fig. 2a). This analysis shows that drag values that are low and comparable to those measured experimentally arise only when the common modes predominate. We next simulated stereociliary motions that matched the experimental records in time resolution and computed the associated coherence, which exceeded 0.95 up to 5 kHz (Fig. 2b). Changing the elastic coupling in the model revealed its importance at low frequencies, whereas viscous coupling intervened at higher frequencies.

These results show that the magnitude of the relative motion in a hair bundle depends on the balance between hydrodynamic and elastic forces. That hair bundles undergoing Brownian motion display a high coherence<sup>11</sup> indicates that the relative mode is very small, which makes it difficult to detect and quantify. We therefore devised an experiment in which hair bundles were stimulated at physiological amplitudes to evoke channel gating and cause intrinsic oscillations at the combination frequencies (Supplementary Information section 7). Because the gating of each mechanotransduction channel in a hair bundle changes



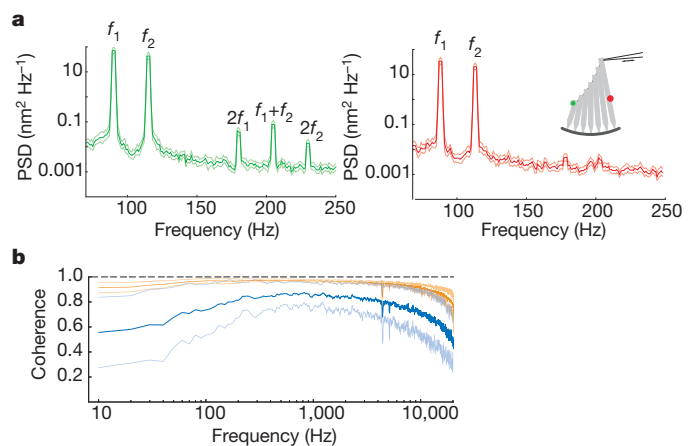
**Figure 2 | Fluid–structure interactions in a stochastic model.** **a**, Calculations for a model with only pivotal stiffness and hydrodynamic drag, two degrees of freedom per stereocilium, and no kinocilium yield 122 eigenmodes, of which four representative examples are shown. The eigenmodes of the damping matrix progress from a collective mode with a low-drag eigenvalue to a relative mode that is a thousand times as dissipative. The reported eigenvalues are expressed in multiples of the smallest one. **b**, The calculated coherence of motion for a hair bundle with a top-connector stiffness of  $20 \text{ mN m}^{-1}$  (orange) or  $20 \mu\text{N m}^{-1}$  (blue) illustrates the importance of elastic linkages at low frequencies and of viscous coupling at high frequencies.

the force in the associated tip link<sup>17</sup>, it must cause a relative motion of the interconnected stereocilia that is balanced by the frictional drag and elastic linkages. Blocking the distortion products at one edge of the hair bundle while measuring the relative motion at the opposite edge allowed us to isolate and quantify the amount of play between adjacent stereocilia during small deflections, assess the forces at play and compare the results with our model.

In agreement with a previous report<sup>18</sup>, using a flexible glass probe attached to a hair bundle's tall edge to stimulate it at two frequencies evoked distortion products at several combination frequencies. These distortion products were robust at both edges of a hair bundle and disappeared when the tip links were disrupted by 1,2-bis(*o*-aminophenoxy) ethane-*N,N,N',N'*-tetraacetic acid (BAPTA), confirming that the distortion was caused by the gating of mechanotransduction channels. We then used a stiff glass probe to stimulate the long edge of the hair bundle. The rigid probe in this key experiment prevented any internally generated motion from contaminating the signal at the tall edge, which therefore consisted purely of the two excitation frequencies. With this constraint, the distortion products were significant only at the free, short edge of the hair bundle (Fig. 3a).

We related the distortion of the short-edge motion to the linear displacement by a power series. The inverse of the quadratic term of this fit was  $0.14 \pm 0.12 \mu\text{m}$  ( $n = 8$ ) for the flexible probe and  $1.6 \pm 0.9 \mu\text{m}$  ( $n = 4$ ) for the stiff probe. The distortions were therefore reduced to less than a tenth of their original value when the bundle's tall edge was forced to follow the stimulus signal exactly. The finite-element model with viscous coupling replicated this effect, with the top-connector stiffness determined independently from the other experimental data. The remaining distortions revealed that the relative movement between adjacent stereocilia was less than a nanometre, only a few times the size of a water molecule (Fig. 3a).

To confirm further the correspondence between experiment and modelling, we tested the prediction that removal of the horizontal top connectors should diminish the coherence (Fig. 1c) and increase the overall drag (Figs 1b and 2a). We placed hair bundles in a  $\text{Ca}^{2+}$ -free, iso-osmotic solution of mannitol having the same viscosity as



**Figure 3 | Experimental verification of model predictions.** **a**, Power spectra reveal that exciting a hair bundle with a stiff glass probe at two frequencies ( $f_1 = 90 \text{ Hz}$  and  $f_2 = 115 \text{ Hz}$ ) generates distortion products marked by peaks of power-spectral density (PSD) at the second harmonics ( $2f_1 = 180 \text{ Hz}$  and  $2f_2 = 230 \text{ Hz}$ ) and at the combination frequency ( $f_1 + f_2 = 205 \text{ Hz}$ ). Because the stiff probe suppresses internally generated movements at the tall edge (right panel), the distortion products are present only at the free short edge (left panel). The presence of distortion products directly demonstrates the relative mode of motion within the array. The schematic diagram of a hair bundle in the inset indicates the stimulating probe attached to the bundle's top and the positions of the red and green laser spots used in the interferometric measurements. **b**, The coherence in perilymph (orange) declines appreciably in the presence of mannitol (blue), which disrupts the horizontal top connectors. The mean values are accompanied by 95% confidence intervals in light orange and light blue, respectively.

saline solution but a lower ionic strength. This medium has been reported to remove the top connectors<sup>19</sup>, and we verified the treatment's effect by transmission electron microscopy (Supplementary Information section 8). After 20 min of treatment, the top connectors were overstretched or broken, but not entirely absent (data not shown). Some elastic coupling thus persisted in mannitol. As our model predicted, the procedure decoupled the stereocilia and increased the drag, although quantitatively the effect was variable from cell to cell, presumably because of heterogeneity in the residual top connectors. For the same six cells in both conditions, the coherence between 100 Hz and 5 kHz declined from  $0.96 \pm 0.01$  in perilymph to  $0.83 \pm 0.12$  in mannitol (Fig. 3b). At the same time, the drag coefficient in mannitol increased to  $99 \pm 63 \text{ nN s m}^{-1}$ . Together with the close match between the coherence values in the experiment and in the models, this and the results above confirm the accuracy of the numerical models and indicate that they capture the essential physics of the fluid–structure interactions in a hair bundle.

In conclusion, because all stereocilia and the liquid between them move in unison over the whole auditory spectrum, with the relative motions apparent only on a sub-nanometre scale, most stereocilia inside the hair bundle are shielded from the external liquid and experience little viscous drag. Although viscous forces might be thought to impair sensitivity and frequency selectivity, the hair bundle's structure actually minimizes energy dissipation, making it easier for the active process to keep the ear tuned. The tight clustering of stereocilia even transforms liquid viscosity into an asset by using it as a simple means of activating numerous mechanosensitive ion channels in concert.

## METHODS SUMMARY

The methods used in this study are described in the Supplementary Information. Force measurements on a scaled hair-bundle model respected the physiological character of the liquid flow. The finite-element method provided approximate solutions to partial differential equations reflecting the hair bundle's geometry. The small amplitudes of motion allowed the elimination of nonlinear terms. The velocity variable of the liquid was replaced with the time derivative of the displacement; fluid pressure was approximated by linear shape functions and the displacements of liquid and solid were approximated by quadratic functions. The hydrodynamic forces between stereocilia were estimated analytically by solving the Stefan–Reynolds equations under the lubrication approximation, which is valid when the gaps between adjacent stereocilia are much smaller than their diameter. Stochastic simulations based on these results were performed for a system of linearly coupled dynamic variables, following a Langevin description with Gaussian white noise at room temperature. The integration procedure was validated by choosing time steps small enough to ensure that the results were independent of the increment. The robustness of our conclusions was investigated by a detailed parameter-variation study. We tested the effects of inertia and of the estimated top-connector stiffness and confirmed the validity of our conclusions for mammalian hair bundles.

Dual-beam differential interferometry was used to record stereociliary motions with sub-nanometre spatial and sub-millisecond temporal resolution. Fourier analysis of the records was performed with the multitaper method to obtain coherence spectra as well as stiffness and drag coefficients. Distortion products were evoked by stimulating hair bundles with calibrated glass probes. These results were used to verify the predictions of the numerical model and to measure the relative mode of motion between stereocilia directly. Transmission and scanning electron microscopy was performed by standard techniques with minor modifications.

Received 5 August 2010; accepted 24 March 2011.

Published online 22 May 2011.

1. Hudspeth, A. J. How the ear's works work. *Nature* **341**, 397–404 (1989).
2. Hudspeth, A. J. Making an effort to listen: mechanical amplification in the ear. *Neuron* **59**, 530–545 (2008).

3. Hudspeth, A. J., Jülicher, F. & Martin, P. A critique of the critical cochlea: Hopf—a bifurcation—is better than none. *J. Neurophysiol.* **104**, 1219–1229 (2010).
4. Stokes, G. G. On the effect of the internal friction of fluids on the motion of pendulums. *Trans. Camb. Phil. Soc.* **9**, 1–86 (1850).
5. Batchelor, G. K. *An Introduction to Fluid Dynamics* 353–364 (Cambridge University Press, 2000).
6. Tamada, K. & Fujikawa, H. The steady two-dimensional flow of viscous liquid at low Reynolds numbers passing through an infinite row of equal parallel circular cylinders. *Q. J. Mech. Appl. Math.* **10**, 425–432 (1957).
7. Yeom, J., Agonafer, D. D., Han, J.-H. & Shannon, M. A. Low Reynolds number flow across an array of cylindrical microposts in a microchannel and figure-of-merit analysis of micropost-filled microreactors. *J. Micromech. Microeng.* **19**, doi:10.1088/0960-1317/19/6/065025 (2009).
8. Denk, W., Webb, W. W. & Hudspeth, A. J. Mechanical properties of sensory hair bundles are reflected in their Brownian motion measured with a laser differential interferometer. *Proc. Natl Acad. Sci. USA* **86**, 5371–5375 (1989).
9. Perrin, F. Mouvement brownien d'un ellipsoïde (I). Dispersion diélectrique pour des molécules ellipsoïdales. *J. Phys. Radium* **5**, 497–511 (1934).
10. Broersma, S. Viscous force and torque constant for a cylinder. *J. Chem. Phys.* **74**, 6989–6990 (1981).
11. Kozlov, A. S., Risler, T. & Hudspeth, A. J. Coherent motion of stereocilia assures the concerted gating of hair-cell transduction channels. *Nature Neurosci.* **10**, 87–92 (2007).
12. Pickles, J. O. A model for the mechanics of the stereociliar bundle on acousticolateral hair cells. *Hear. Res.* **68**, 159–172 (1993).
13. Cotton, J. & Grant, W. Computational models of hair cell bundle mechanics: II. Simplified bundle models. *Hear. Res.* **197**, 105–111 (2004).
14. Karavitsaki, K. D. & Corey, D. P. Sliding adhesion confers coherent motion to hair cell stereocilia and parallel gating to transduction channels. *J. Neurosci.* **30**, 9051–9063 (2010).
15. Martin, P., Mehta, A. D. & Hudspeth, A. J. Negative hair-bundle stiffness betrays a mechanism for mechanical amplification by the hair cell. *Proc. Natl Acad. Sci. USA* **97**, 12026–12031 (2000).
16. Verpy, E. *et al.* Stereocilin connects outer hair cells stereocilia to one another and to the tectorial membrane. *J. Comp. Neurol.* **519**, 194–210 (2011).
17. Howard, J. & Hudspeth, A. J. Compliance of the hair bundle associated with gating of mechano-electrical transduction channels in the bullfrog's saccular hair cell. *Neuron* **1**, 189–199 (1988).
18. Jaramillo, F., Markin, V. S. & Hudspeth, A. J. Auditory illusions and the single hair cell. *Nature* **364**, 527–529 (1993).
19. Neugebauer, D. C. & Thurm, U. Surface charges of the membrane and cell adhesion substances determine the structural integrity of hair bundles from the inner ear of fish. *Cell Tissue Res.* **249**, 199–207 (1987).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank A. J. Hinterwirth for assistance in constructing the interferometer and B. Fabella for programming the experimental software; M. Fleischer for help with programming the fluid finite-element model; R. Gärtner and A. Voigt for discussions of the finite-element model and stochastic computations; M. Lenz for discussions of stochastic computations and the analytic derivation of fluid-mediated interactions; and O. Ahmad, D. Andor and M. O. Magnasco for discussions about data analysis. This research was funded by National Institutes of Health grant DC000241. Computational resources were provided by the Center for Information Services and High Performance Computing of the Technische Universität Dresden. J.B. was supported by grants Gr 1388/14 and Vo 899/6 from the Deutsche Forschungsgemeinschaft. A.S.K. was supported by the Howard Hughes Medical Institute, of which A.J.H. is an Investigator.

**Author Contributions** A.S.K. organized the collaboration, designed and performed the experiments, analysed data, and wrote most of the manuscript. J.B. developed the finite-element formulation and conducted the corresponding computations, implemented the stochastic modelling, derived analytic estimates of fluid-mediated interactions, wrote the corresponding Supplementary Information sections, analysed data and edited the manuscript. T.R. derived analytic estimates of fluid-mediated interactions, developed the stochastic models, implemented the data analysis, wrote the corresponding Supplementary Information sections and edited the manuscript. C.P.C.V. built the scaled model and performed the corresponding experiment. A.J.H. designed the experiments, performed the electron microscopy and edited the manuscript.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at [www.nature.com/nature](http://www.nature.com/nature). Correspondence and requests for materials should be addressed to A.J.H. ([hudspaj@rockefeller.edu](mailto:hudspaj@rockefeller.edu)).



# Transcriptomic analysis of autistic brain reveals convergent molecular pathology

Irina Voineagu<sup>1</sup>, Xinchun Wang<sup>2</sup>, Patrick Johnston<sup>3</sup>, Jennifer K. Lowe<sup>1</sup>, Yuan Tian<sup>1</sup>, Steve Horvath<sup>4</sup>, Jonathan Mill<sup>3</sup>, Rita M. Cantor<sup>4</sup>, Benjamin J. Blencowe<sup>2</sup> & Daniel H. Geschwind<sup>1,4</sup>

Autism spectrum disorder (ASD) is a common, highly heritable neurodevelopmental condition characterized by marked genetic heterogeneity<sup>1–3</sup>. Thus, a fundamental question is whether autism represents an aetiologically heterogeneous disorder in which the myriad genetic or environmental risk factors perturb common underlying molecular pathways in the brain<sup>4</sup>. Here, we demonstrate consistent differences in transcriptome organization between autistic and normal brain by gene co-expression network analysis. Remarkably, regional patterns of gene expression that typically distinguish frontal and temporal cortex are significantly attenuated in the ASD brain, suggesting abnormalities in cortical patterning. We further identify discrete modules of co-expressed genes associated with autism: a neuronal module enriched for known autism susceptibility genes, including the neuronal specific splicing factor *A2BP1* (also known as *FOX1*), and a module enriched for immune genes and glial markers. Using high-throughput RNA sequencing we demonstrate dysregulated splicing of *A2BP1*-dependent alternative exons in the ASD brain. Moreover, using a published autism genome-wide association study (GWAS) data set, we show that the neuronal module is enriched for genetically associated variants, providing independent support for the causal involvement of these genes in autism. In contrast, the immune-glial module showed no enrichment for autism GWAS signals, indicating a non-genetic aetiology for this process. Collectively, our results provide strong evidence for convergent molecular abnormalities in ASD, and implicate transcriptional and splicing dysregulation as underlying mechanisms of neuronal dysfunction in this disorder.

We analysed post-mortem brain tissue samples from 19 autism cases and 17 controls from the Autism Tissue Project and the Harvard brain bank (Supplementary Table 1) using Illumina microarrays. For each individual, we profiled three regions previously implicated in autism<sup>5</sup>: superior temporal gyrus (STG, also known as Brodmann's area (BA) 41/42), prefrontal cortex (BA9) and cerebellar vermis. After filtering for high-quality array data (Methods), we retained 58 cortex samples (29 autism, 29 controls) and 21 cerebellum samples (11 autism, 10 controls) for further analysis (see Methods for detailed sample description). We identified 444 genes showing significant expression changes in autism cortex samples (DS1, Fig. 1b), and only 2 genes were differentially expressed between the autism and control groups in cerebellum (Methods), indicating that gene expression changes associated with autism were more pronounced in the cerebral cortex, which became the focus of further analysis (Supplementary Table 2). There was no significant difference in age, post mortem interval (PMI), or RNA integrity numbers (RIN) between autism and control cortex samples (Supplementary Fig. 1, Methods).

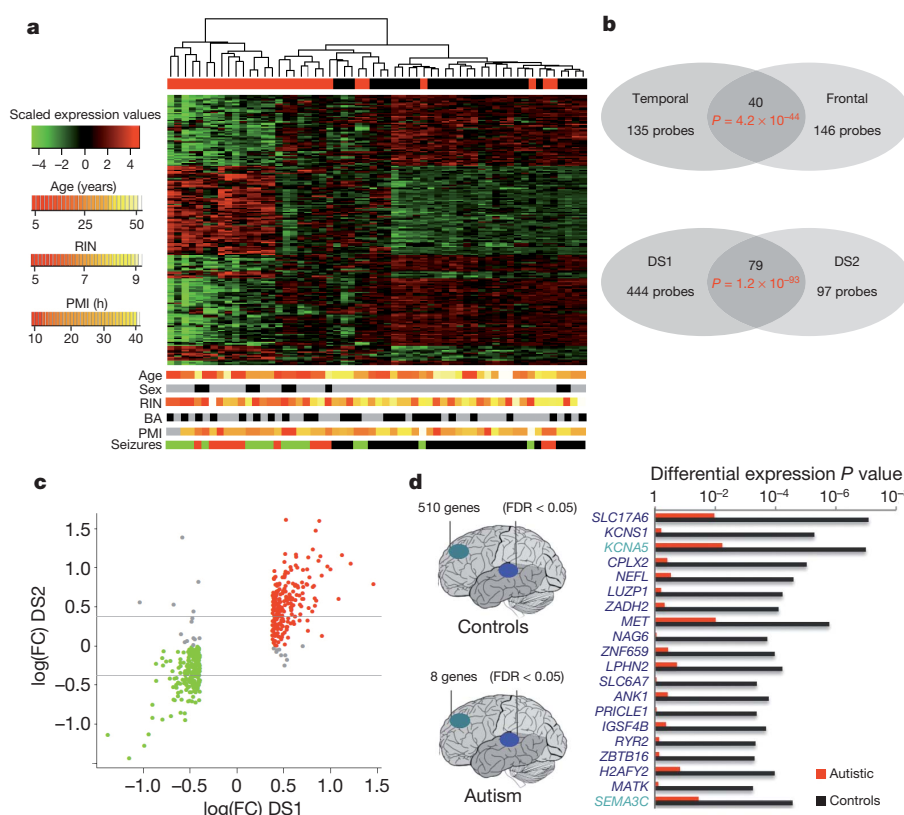
Supervised hierarchical clustering based on the top 200 differentially expressed genes showed distinct clustering of the majority of autism cortex samples (Fig. 1a), including one case that was simultaneously

found to have a 15q duplication (Methods, Supplementary Table 1), which is known to cause 1% of ASD<sup>6</sup>. Cortex samples from ten of the cases coalesced in a single tight-clustering branch of the dendrogram. Clustering was independent of age, sex, RIN, PMI, co-morbidity of seizures, or medication (Fig. 1a and Supplementary Fig. 2c). It is interesting to note that the two ASD cases that cluster with controls (Fig. 1a) are the least severe cases, as assessed by global functioning (Supplementary Table 12). We observed a highly significant overlap between differentially expressed genes in frontal and temporal cortex ( $P = 10^{-44}$ ; Fig. 1b), supporting the robustness of the data and indicating that the autism-specific expression changes are consistent across these cortical areas. We also validated a cross section of the differentially expressed genes by quantitative reverse transcription PCR (RT-PCR) and confirmed microarray-predicted changes in 83% of the genes tested (Methods, Supplementary Fig. 2b). Gene ontology enrichment analysis (Methods) showed that the 209 genes downregulated in autistic cortex were enriched for gene ontology categories related to synaptic function, whereas the upregulated genes ( $N = 235$ ) showed enrichment for gene ontology categories implicated in immune and inflammatory response (Supplementary Table 3).

To test whether these findings were replicable, and to further validate the results in an independent data set, we obtained tissue from an additional frontal cortex region (BA44/45) from nine ASD cases and five controls (DS2; Supplementary Table 4). Three of the cases and all of the controls used for validation were independent from our initial cohort. Ninety-seven genes were differentially expressed in BA44/45 in DS2, and 81 of these were also differentially expressed in our initial cohort ( $P = 1.2 \times 10^{-93}$ , hypergeometric test; Fig. 1b, c). Remarkably, the direction of expression differences between autism and controls was the same as in the initial cohort for all but 2 of the 81 overlapping differentially expressed probes. Hierarchical clustering of DS2 samples based on either the top 200 genes differentially expressed in the initial cohort or the 81 overlapping genes showed distinct separation of cases from controls (Supplementary Fig. 6). In addition, comparison of these differentially expressed results with another, smaller study of the STG in ASD<sup>7</sup>, revealed significant consistency at the level of differentially expressed genes, including downregulation of *DLX1* and *AHI1* (Supplementary Table 5). Thus, differential expression analysis produced robust and highly reproducible results, warranting further refined analysis.

We next applied weighted-gene co-expression network analysis (WGCNA)<sup>8,9</sup> to integrate the expression differences observed between autistic and control cerebral cortex into a higher order, systems level context. We first asked whether there are global differences in the organization of the brain transcriptome between autistic and control brain by constructing separate co-expression networks for the autism and control groups (Methods). The control brain network showed high similarity with the previously described human brain co-expression networks (Supplementary Table 7), consistent with the existence of

<sup>1</sup>Program in Neurogenetics and Neurobehavioral Genetics, Department of Neurology and Semel Institute, David Geffen School of Medicine, University of California, Los Angeles, California 90095-1769, USA. <sup>2</sup>Banting and Best Department of Medical Research, Donnelly Centre, University of Toronto, Toronto, Ontario M5G 1L6, Canada. <sup>3</sup>Institute of Psychiatry, King's College London, London SE5 8AF, UK. <sup>4</sup>Department of Human Genetics, University of California Los Angeles, Los Angeles, California 90095, USA.



**Figure 1 | Gene expression changes in autism cerebral cortex** **a**, Heat map of top 200 genes differentially expressed between autism and control cortex samples. Scaled expression values are colour-coded according to the legend on the left. The dendrogram depicts hierarchical clustering based on the top 200 differentially expressed genes. The top bar (A/C) indicates the disease status: red, autism; black, control. The bottom bars show additional variables for each sample: sex (grey, male; black, female), brain area (black, temporal; grey, frontal), co-morbidity of seizures (green, autism case with seizure disorder; red, autism case without seizure disorder; black, control), age, RNA integrity number (RIN) and post mortem interval (PMI). BA, Brodmann's area. The corresponding scale for quantitative variables is shown on the left. **b**, Top, Venn diagram depicting the overlap between genes differentially expressed in frontal and temporal cortex. Bottom, Venn diagram describing the overlap between genes differentially expressed in the initial cohort (DS1) and the replication cohort (DS2). Differential expression in the initial cohort was assessed at an FDR < 0.05 and fold change > 1.3. The statistical criteria were relaxed to

$P < 0.05$  for the replication data set because it involved fewer samples.

**c**, Expression fold changes for all genes differentially expressed in the initial cohort are plotted on the x-axis against the fold changes for the same genes in the replication cohort on the y-axis. Green, genes downregulated in the autism group in both data sets; red, genes upregulated in the autism group in both data sets; grey, genes with opposite direction of variation in the two data sets. Horizontal lines show fold change threshold for significance. **d**, Diagram depicting the number of genes showing significant expression differences between frontal and temporal cortex in control samples (top) and autism samples (bottom) at FDR < 0.05 (left). The top 20 genes differentially expressed between frontal and temporal cortex in control samples (right). All of the genes shown are also differentially expressed between frontal and temporal cortex in fetal midgestation brain<sup>10</sup>, but show no significant expression differences between frontal and temporal cortex in autism. The horizontal bars depict *P* values for differential expression between frontal and temporal cortex in the autism and control groups.

robust modules of co-expressed genes related to specific cell types and biological functions<sup>8</sup>. Similarly, the majority (87%) of the autism modules showed significant overlap with the previously described human brain modules (Supplementary Table 6), indicating that many features reflecting the general organization of the autism brain transcriptome are consistent with that of the normal human brain.

The expression levels of each module were summarized by the first principal component (the module eigengene), and were used to assess whether modules are related to clinical phenotypes or other experimental variables, such as brain region. Two of the control module eigengenes (cM6, cM13) showed significant differences ( $P < 0.05$ ) between the two cortical regions as expected, whereas none of the ASD modules showed any differences between frontal and temporal cortex. This led us to explore the hypothesis that the normal molecular distinctions between the two cortical regions tested were altered in ASD compared with controls. Remarkably, whereas 174 genes were differentially expressed between control BA9 and BA41 (false discovery rate (FDR) < 1%), none of the genes were differentially expressed in the same regional comparison among the ASD cases. This was not simply an issue of statistical thresholds, as relaxing the statistical criteria for differential expression to an FDR of 5% identified over 500 differentially

expressed genes in controls, and only 8 in ASD brains, confirming the large difference observed in regional cortical differential gene expression between ASD cases and controls (Fig. 1d, Methods). Analysis of differential expression from a data set<sup>10</sup> of gene expression in developing fetal human brain showed a highly significant ( $P = 5.8 \times 10^{-9}$ ) overlap of differentially expressed genes with those found in controls in this study, independently confirming that these genes differentiate normal temporal and frontal lobes. We evaluated the homogeneity of gene expression variance across the autism and control groups using Bartlett's test (Methods) which indicated that increased variance was not the major factor responsible for the striking difference in regional gene expression between ASD and controls (Supplementary Fig. 7 and Supplementary Data).

These data suggest that typical regional differences, many of which are observed during fetal development<sup>10</sup>, are attenuated in frontal and temporal lobe in autism brain, pointing to abnormal developmental patterning as a potential pathophysiological driver in ASD. This is especially interesting in light of a recent anatomical study of five cases with adult autism which demonstrated a reduction in typical ultrastructural differences between three frontal cortical regions in autism<sup>11</sup>. Together, these independent studies provide both molecular and

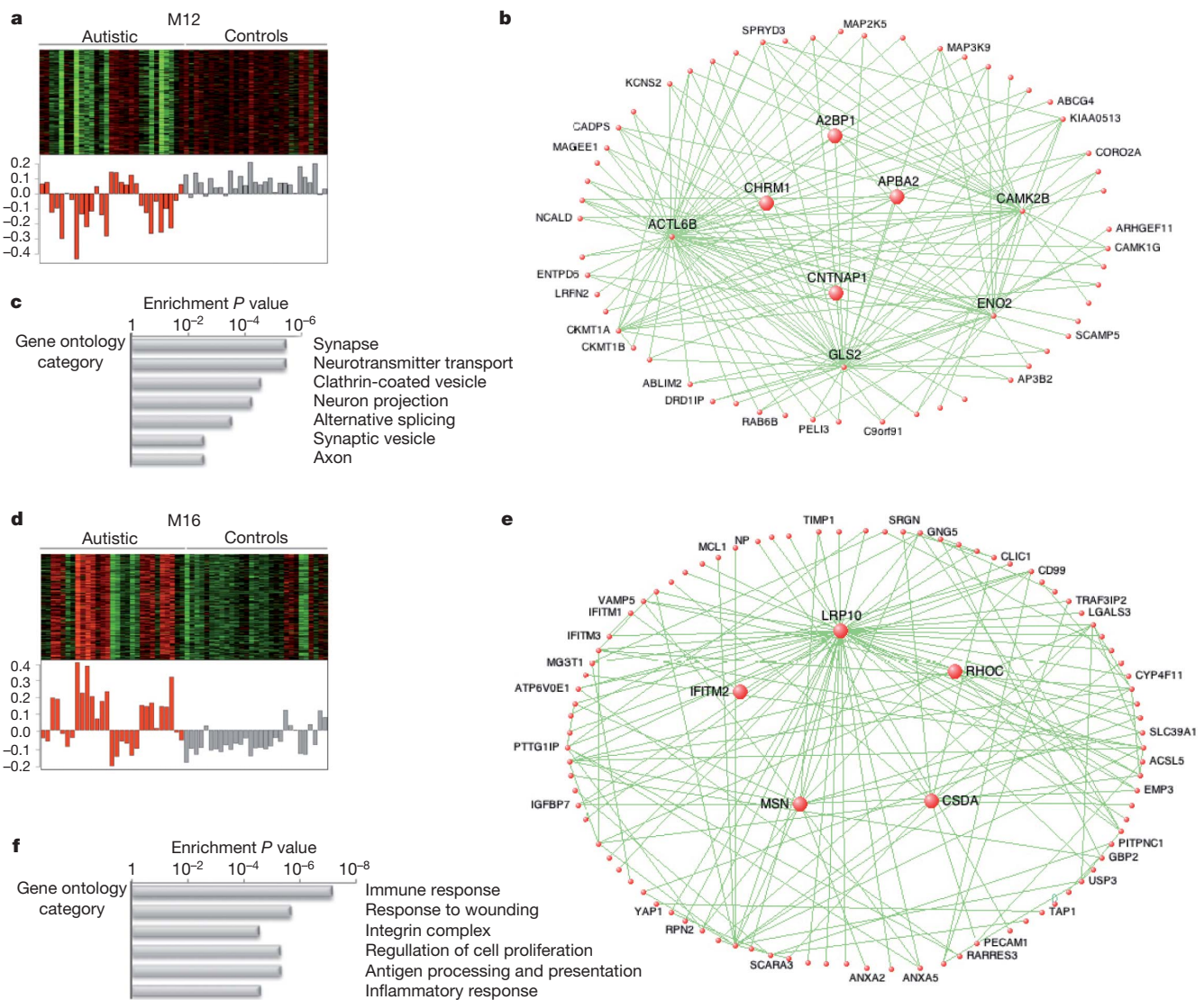
structural evidence suggesting a relative diminution of cortical regional identity in autism.

To identify discrete groups of co-expressed genes showing transcriptional differences between autism and controls, we constructed a co-expression network using the entire data set, composed of both autism and control samples (Methods). As previously shown for complex diseases<sup>12,13</sup> co-expression networks allow analysis of gene expression variation related to multiple disease-related and genetic traits. We assessed module eigengene relationship to autism disease status, age, gender, cause of death, co-morbidity of seizures, family history of psychiatric disease, and medication, providing a complementary assessment of these potential confounders to that performed in the standard differential expression analysis (Supplementary Table 9).

The comparison between autism and control groups revealed two network modules whose eigengenes were highly correlated with disease status, and not any of the potential confounding variables (Supplementary Table 9). We found that the top module (M12) showed highly significant enrichment for neuronal markers (Supplementary Table 9), and high overlap with two neuronal modules previously identified as part of the human brain transcriptional network<sup>8</sup>: a *PVALB*+

interneuron module and a module of genes involved in synaptic function. The M12 eigengene was under-expressed in autism cases, indicating that genes in this module were downregulated in the autistic brain (Fig. 2). Consistent with the pathways identified to be downregulated in autism by differential expression analysis (Supplementary Table 3), the functional enrichment of M12 included the gene ontology categories involved in synaptic function, vesicular transport and neuronal projection.

Remarkably, unlike differentially expressed genes, M12 showed significant overrepresentation of known autism susceptibility genes<sup>2</sup> (Supplementary Table 10;  $P = 6.1 \times 10^{-4}$ ), including *CADPS2*, *AHI1*, *CNTNAP2*, and *SLC25A12*, supporting the increased power of the network-based approach to identify disease-relevant transcriptional changes. A further advantage of network analysis over standard analysis of differential expression is that it allows one to infer the functional relevance of genes based on their network position<sup>9</sup>. The hubs of M12, that is, the genes with the highest rank of M12 membership<sup>8</sup>, were *A2BP1*, *APBA2*, *SCAMP5*, *CNTNAP1*, *KLC2*, and *CHRM1* (Supplementary Data). The first three of these genes have previously been implicated in autism<sup>14–16</sup>, whereas the fourth is a homologue of



**Figure 2 | Gene co-expression modules associated with autism** **a, d**, Heat map of genes belonging to the co-expression module (top). Corresponding module eigengene values (y-axis) across samples (x-axis) (bottom). Red, autism; grey, controls. **b, e**, Visualization of the M12 and M16 modules,

respectively. The top 150 connections are shown for each module. Genes with the highest correlation with the module eigengene value (that is, intramodular hubs) are shown in larger size. **c, f**, Relevant gene ontology categories enriched in the M12 and M16 modules.



the autism susceptibility gene *CNTNAP2* (ref. 17). We highlight the group of genes most strongly connected to the known ASD genes (Supplementary Fig. 5) and emphasize the downregulation of several interneuron markers, such as *DLX1* and *PVALB*, as candidates for future genetic and pathologic investigations.

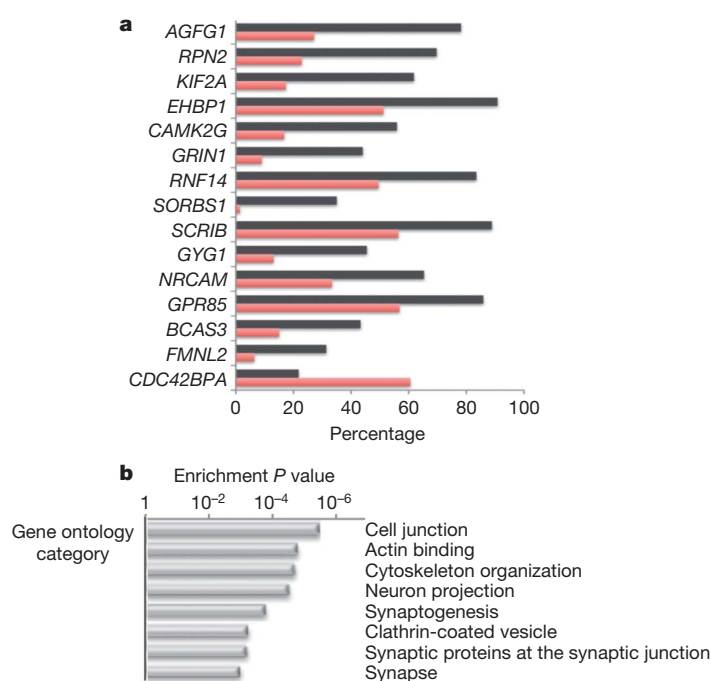
The second module of co-expressed genes highly related to autism disease status, M16, was enriched for astrocyte markers and markers of activated microglia (Supplementary Table 9), as well as for genes belonging to immune and inflammatory gene ontology categories (Fig. 2). This module, which was upregulated in ASD brain, showed significant similarity to two modules identified in previous studies of normal human brain<sup>8</sup>: an astrocyte module and a microglial module. Consistent with this functional annotation, two of the hubs of the M16 module were known astrocyte markers (*ADFP*, also known as *PLIN2*, and *IFITM2*).

One of the hubs of the M12 module was *A2BP1*, a neural- and muscle-specific alternative splicing regulator<sup>18</sup> and the only splicing factor previously implicated in ASD<sup>16</sup>. Because *A2BP1* was downregulated in several ASD cases (Supplementary Fig. 8), this observation provided a unique opportunity to identify potential disease-relevant *A2BP1* targets. Whereas *A2BP1*-regulated alternative exons have been predicted genome-wide<sup>19</sup>, few genes have been experimentally validated as *A2BP1* targets<sup>20</sup>. To identify potential *A2BP1*-dependent differential splicing events in ASD brain, we performed high-throughput RNA sequencing (RNA-Seq) on three autism samples with significant downregulation of *A2BP1* (average fold change by quantitative RT-PCR = 5.9) and three control samples with average *A2BP1* levels. We identified 212 significant alternative splicing events (Supplementary Data). Among these, 36 had been defined<sup>19</sup> as predicted targets of *A2BP1*/2, which represents a highly significant overlap (36/176;  $P = 2.2 \times 10^{-16}$ ). In addition, five previously validated *A2BP1* targets showed evidence of alternative splicing, four of which (*ATP5C1*, *ATP2B1*, *GRIN1* and *MEF2C*) were confirmed as having differential splicing between ASD samples with low *A2BP1* expression and control samples, indicating that we were able to identify a high proportion of the expected *A2BP1*-dependent differential splicing events. We also observe that alternative exons with increased skipping in ASD relative to control cases are significantly enriched for *A2BP1* motifs in adjacent, downstream intronic sequences ( $P = 1.09 \times 10^{-7}$ , Fisher's exact test), consistent with previous data<sup>19</sup>.

The top gene ontology categories enriched among ASD differential splicing genes highly overlapped with the gene ontology categories found to be enriched in the M12 module (Fig. 3b). In addition, *A2BP1* target genes showed enrichment for actin-binding proteins and genes involved in cytoskeleton reorganization (Fig. 3b). Among top predicted *A2BP1*-dependent differential splicing events (Fig. 3a) are *CAMK2G*, which also belongs to the M12 module, as well as *NRCAM* and *GRIN1*. The latter are proteins involved in synaptogenesis, in which allelic variants have been associated with autism and schizophrenia, respectively<sup>21,22</sup>.

RT-PCR assays confirmed a high proportion (85%) of the tested differential splicing changes involving predicted *A2BP1* targets (Supplementary Fig. 8). We further tested the differential splicing events validated by RT-PCR in three independent ASD cases with decreased *A2BP1* levels and confirmed the predicted changes in alternative splicing (Supplementary Fig. 8), indicating that the observed differential splicing events are indeed associated with reduced *A2BP1* levels, rather than due to inter-individual variability. The RNA-Seq data thus provides validation of the functional groups of genes identified by co-expression analysis, and evidence for a convergence of transcriptional and alternative-splicing abnormalities in the synaptic and signalling pathogenesis of ASD.

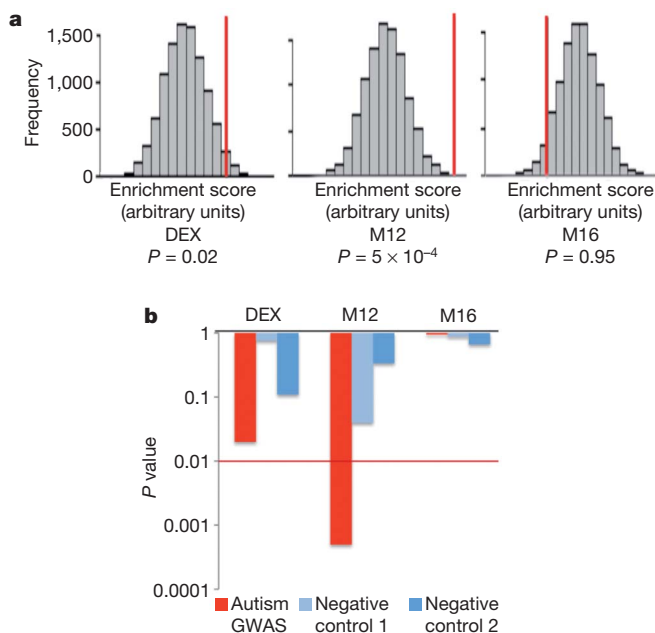
To test whether our findings are more generalizable, and determine whether the autism-associated transcriptional differences observed are likely to be causal, versus collateral effects or environmentally-induced changes, we tested whether our co-expression modules or



**Figure 3 | A2BP1-dependent differential splicing events** **a**, Top *A2BP1*-specific differential splicing events. Differential splicing events showing the most significant differences in alternative splicing between low-*A2BP1* autism cases and controls as well as differential splicing differences consistent with the *A2BP1* binding site position. The horizontal axis depicts the percentage of transcripts including the alternative exon. Red, autism samples; black, control samples. **b**, Relevant gene ontology categories enriched in the set of genes containing exons differentially spliced between low-*A2BP1* autism cases and controls.

the differentially expressed genes show enrichment for autism genetic association signals. M12 showed highly significant enrichment for association signals ( $P = 5 \times 10^{-4}$ ), but neither M16 nor the list of differentially expressed genes showed such enrichment (Fig. 4). As a negative control, we performed the same set-enrichment analysis using two GWAS studies for non-psychiatric disease performed on the same genotyping platform: a genome-wide association for hair colour<sup>23</sup>, and a GWAS study of warfarin maintenance dose<sup>24</sup> finding no significant enrichment of the association signal (Fig. 4b, Supplementary Fig. 4). These results indicate that (1) M12 consists of a set of genes that are supported by independent lines of evidence to be causally involved in ASD pathophysiology, and (2) the upregulation of immune response genes in the autistic brain observed by us and others<sup>25</sup> has no evidence of a common genetic component.

Our system-level analysis of the ASD brain transcriptome demonstrates the existence of convergent molecular abnormalities in ASD for the first time, providing a molecular neuropathological basis for the disease, whose genetic, epigenetic, or environmental aetiologies can now be directly explored. The genome-wide analysis performed here significantly extends previous findings implicating synaptic dysfunction, as well as microglial and immune dysregulation in ASD<sup>6</sup> by providing an unbiased systematic assessment of transcriptional alterations and their genetic basis. We show that the transcriptome changes observed in ASD brain converge with GWAS data in supporting the genetic basis of synaptic and neuronal signalling dysfunction in ASD, whereas immune changes have a less pronounced genetic component and thus are most likely either secondary phenomena or caused by environmental factors. Because immune molecules and cells such as microglia have a role in synaptic development and function<sup>26</sup>, we speculate that the observed immune upregulation may be related to abnormal ongoing plasticity in the ASD brain. The striking attenuation of gene expression differences observed here between frontal and temporal cortex in ASD is likely to represent a defect of developmental patterning and provides a strong rationale for further studies to assess the pervasiveness



**Figure 4 | GWAS set enrichment analysis** **a**, GWAS set enrichment analysis using the discovery AGRE cohort from ref. 27. For each gene set (DEX, differentially expressed genes; M12 and M16) the null distribution of the enrichment score generated by 10,000 random permutations is shown (x-axis) and the enrichment score for the gene set is depicted by a red vertical line. A *P* value <0.01 was considered significant to correct for multiple comparisons. **b**, GWAS signal enrichment of differentially expressed genes and the autism-associated co-expression modules M12 and M16. Enrichment *P* values are shown for an autism GWAS data set (ref. 27, AGRE discovery cohort) as well as two control data sets consisting of GWAS studies of non-psychiatric traits: ref. 23 (Negative control 1) and ref. 24 (Negative control 2). The red line marks the *P* value threshold for significance.

of transcriptional patterning abnormalities across the ASD brain. We also demonstrate for the first time alterations in differential splicing associated with *A2BP1* levels in the ASD brain, and show that many of the affected exons belong to genes involved in synaptic function. Finally, given current evidence of genetic overlap between ASD and other neurodevelopmental disorders including schizophrenia and attention deficit hyperactivity disorder (ADHD), the data provide a new pathway-based framework from which to assess the enrichment of genetic association signals in other allied psychiatric disorders.

## METHODS SUMMARY

**Brain tissue.** Post-mortem brain tissue was obtained from the Autism Tissue Project and the Harvard Brain Bank as well as the MRC London Brain bank for Neurodegenerative Disease. Detailed information on the autism cases included in this study is available in Methods.

**Microarrays and RNA-seq.** Total RNA was extracted from 100 mg of tissue using a Qiagen miRNA kit according to the manufacturer's protocol. Expression profiles were obtained using Illumina Ref8 v3 microarrays. RNA-seq was performed on the Illumina GAIIX, as per the manufacturer's instructions. Further detailed information on data analysis is available in Methods.

Full detailed Methods accompany this paper as Supplementary Information.

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

Received 12 December 2010; accepted 13 April 2011.

Published online 25 May 2011.

1. Durand, C. M. *et al.* Mutations in the gene encoding the synaptic scaffolding protein SHANK3 are associated with autism spectrum disorders. *Nature Genet.* **39**, 25–27 (2006).
2. Pinto, D. *et al.* Functional impact of global rare copy number variation in autism spectrum disorders. *Nature* **466**, 368–372 (2010).
3. Sebat, J. *et al.* Strong association of de novo copy number mutations with autism. *Science* **316**, 445–449 (2007).

4. Geschwind, D. H. Autism: many genes, common pathways? *Cell* **135**, 391–395 (2008).
5. Amaral, D. G., Schumann, C. M. & Nordahl, C. W. Neuroanatomy of autism. *Trends Neurosci.* **31**, 137–145 (2008).
6. Abrahams, B. S. & Geschwind, D. H. Advances in autism genetics: on the threshold of a new neurobiology. *Nature Rev. Genet.* **9**, 341–355 (2008).
7. Garbett, K. *et al.* Immune transcriptome alterations in the temporal cortex of subjects with autism. *Neurobiol. Dis.* **30**, 303–311 (2008).
8. Oldham, M. C. *et al.* Functional organization of the transcriptome in human brain. *Nature Neurosci.* **11**, 1271–1282 (2008).
9. Zhang, B. & Horvath, S. A general framework for weighted gene co-expression network analysis. *Stat. Appl. Genet. Mol. Biol.* **4**, 17 (2005).
10. Johnson, M. B. *et al.* Functional and evolutionary insights into human brain development through global transcriptome analysis. *Neuron* **62**, 494–509 (2009).
11. Zikopoulos, B. & Barbas, H. Changes in prefrontal axons may disrupt the network in autism. *J. Neurosci.* **30**, 14595–14609 (2010).
12. Chen, Y. *et al.* Variations in DNA elucidate molecular networks that cause disease. *Nature* **452**, 429–435 (2008).
13. Plaisier, C. L. *et al.* A systems genetics approach implicates *USF1*, *FADS3*, and other causal candidate genes for familial combined hyperlipidemia. *PLoS Genet.* **5**, e1000642 (2009).
14. Babatz, T. D., Kumar, R. A., Sudi, J., Dobyns, W. B. & Christian, S. L. Copy number and sequence variants implicate *APBA2* as an autism candidate gene. *Autism Res.* **2**, 359–364 (2009).
15. Castermans, D. *et al.* *SCAMP5*, *NBEA* and *AMISYN*: three candidate genes for autism involved in secretion of large dense-core vesicles. *Hum. Mol. Genet.* **19**, 1368–1378 (2010).
16. Martin, C. L. *et al.* Cytogenetic and molecular characterization of *A2BP1/FOX1* as a candidate gene for autism. *Am. J. Med. Genet. B. Neuropsychiatr. Genet.* **144B**, 869–876 (2007).
17. Alarcón, M. *et al.* Linkage, association, and gene-expression analyses identify *CNTNAP2* as an autism-susceptibility gene. *Am. J. Hum. Genet.* **82**, 150–159 (2008).
18. Underwood, J. G., Boutz, P. L., Dougherty, J. D., Stoilov, P. & Black, D. L. Homologues of the *Caenorhabditis elegans* Fox-1 protein are neuronal splicing regulators in mammals. *Mol. Cell. Biol.* **25**, 10005–10016 (2005).
19. Zhang, C. *et al.* Defining the regulatory network of the tissue-specific splicing factors Fox-1 and Fox-2. *Genes Dev.* **22**, 2550–2563 (2008).
20. Lee, J. A., Tang, Z. Z. & Black, D. L. An inducible change in Fox-1/A2BP1 splicing modulates the alternative splicing of downstream neuronal target exons. *Genes Dev.* **23**, 2284–2293 (2009).
21. Moy, S. S., Nonneman, R. J., Young, N. B., Demyanenko, G. P. & Maness, P. F. Impaired sociability and cognitive function in *Nrcam*-null mice. *Behav. Brain Res.* **205**, 123–131 (2009).
22. Zhao, X. *et al.* Significant association between the genetic variations in the 5' end of the N-methyl-D-aspartate receptor subunit gene *GRIN1* and schizophrenia. *Biol. Psychiatry* **59**, 747–753 (2006).
23. Han, J. *et al.* A genome-wide association study identifies novel alleles associated with hair color and skin pigmentation. *PLoS Genet.* **4**, e1000074 (2008).
24. Cooper, G. M. *et al.* A genome-wide scan for common genetic variants with a large influence on warfarin maintenance dose. *Blood* **112**, 1022–1027 (2008).
25. Morgan, J. T. *et al.* Microglial activation and increased microglial density observed in the dorsolateral prefrontal cortex in autism. *Biol. Psychiatry* **68**, 368–376 (2010).
26. Boulanger, L. M. Immune proteins in brain development and synaptic plasticity. *Neuron* **64**, 93–109 (2009).
27. Wang, K. *et al.* Common genetic variants on 5p14.1 associate with autism spectrum disorders. *Nature* **459**, 528–533 (2009).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We are grateful for the efforts of the Autism Tissue Program (ATP) of Autism Speaks and the families that have enrolled in the ATP, which made this work possible. We also thank S. Scherer and R. Wintle for sharing their SNP genotyping data on the AGP samples with us before its publication. We would also like to thank B. Abrahams for help in the initial stages of the project, B. Fogel, G. Konopka, N. Barbosa-Morais and J. Bomar for critically reading the manuscript, M. Lazaro for help with tissue dissection, and C. Vijayendran and K. Winden for useful discussions. This work was funded by an Autism Center of Excellence Network Grant from NIMH 5R01MH081754-03 and NIMH R37MH060233 to D.H.G. and by grants from the Canadian Institutes of Health Research and Genome Canada through the Ontario Genomics Institute to B.J.B. and others.

**Author Contributions** I.V. and D.H.G. designed the study and wrote the manuscript. I.V. performed experiments, analysed the data and conducted the GWAS set enrichment analysis. X.W. and B.J.B. analysed the RNA sequencing data. J.K.L. contributed to the GWAS set enrichment analysis. Y.T. performed some of the microarray qRT-PCR validation experiments. R.M.C. supervised the GWAS set enrichment analysis. S.H. supervised the WGCNA analysis. P.J. and J.M. provided dissected tissue for the replication experiment. All authors discussed the results and commented on the manuscript.

**Author Information** All microarray and RNA-seq data are deposited in GEO under accession number GSE28521. Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at [www.nature.com/nature](http://www.nature.com/nature). Correspondence and requests for materials should be addressed to D.H.G. ([dhg@mednet.ucla.edu](mailto:dhg@mednet.ucla.edu)).

## METHODS

**Brain tissue samples.** Brain tissue samples from 19 autism cases and 17 controls were obtained from the Autism Tissue Project (ATP) and the Harvard Brain Bank. For each brain, tissue was obtained from frontal cortex (BA9), temporal cortex (BA41/42 or BA22) and cerebellum (vermis), with the exception of three controls lacking the cerebellum sample (Supplementary Table 1). For the replication experiment, frontal cortex tissue (BA44/45) from nine ASD cases and five controls were obtained from the ATP and MRC London Brain bank for Neurodegenerative Disease respectively (Supplementary Table 4).

For all of the autism cases, clinical information is available upon request from ATP (<http://www.autismtissueprogram.org>), including the ADI-R diagnostic scores. Supplementary Table 12 contains a summary of clinical characteristics. Although autism cases with known genetic causes were not included in this study, one case with a chromosome 15q duplication was identified for AN17138 by high density small nucleotide polymorphism (SNP) arrays<sup>28</sup> during the course of this study. The ATP cases were genotyped with high-density SNP arrays and with two exceptions all are Caucasians. The two Asian samples cluster with the other ASD cases in the current study, and are not distinguishable from the Caucasian cases based on clustering by gene expression.

**RNA extractions and microarrays.** Total RNA was extracted from approximately 100 mg of frozen tissue, using the Qiagen miRNA kit. RNA concentration was assessed by a NanoDrop spectrophotometer and RNA quality was measured using an Agilent Bioanalyzer. All RNA samples included in the expression analysis had an RNA integrity number (RIN) > 5. cDNA labelling and hybridizations on Illumina Ref8 v3 microarrays were performed according to the manufacturer's protocol.

**Microarray data analysis.** Microarray data analysis was performed using the R software and Bioconductor packages. Raw expression data were log<sub>2</sub> transformed and normalized by quantile normalization. Data quality control criteria included high inter-array correlation (Pearson correlation coefficients > 0.85) and detection of outlier arrays based on mean inter-array correlation and hierarchical clustering. Probes were considered robustly expressed if the detection *P* value was < 0.05 for at least half of the samples in the data set. Cortex samples (58: 29 autism, 29 controls) and cerebellum samples (21: 11 autism, 10 controls) fulfilled all data quality control criteria. The 29 autism cortex samples included tissue from 13 ASD cases with both frontal and temporal cortex and 3 ASD cases with frontal cortex only (in total 16 frontal cortex and 13 temporal cortex ASD samples). The 29 autism control samples also included tissue from 13 controls with both frontal and temporal cortex and 3 controls with frontal cortex only (in total 16 frontal cortex and 13 temporal cortex control samples).

Initially, all samples were normalized together to assess clustering by brain region. As expected, we observed distinct clustering of cortex and cerebellum samples (Supplementary Fig. 2A). For subsequent analyses, cortex samples and cerebellum samples were normalized and analysed separately.

**Differential expression.** Differential expression was assessed using the SAM package (significance analysis of microarrays, <http://www-stat.stanford.edu/~tibs/SAM>) and unless otherwise specified the significance threshold was FDR < 0.05 and fold changes > 1.3. Given that SAM is less sensitive in detecting differentially expressed genes for small number of samples, for the replication cohort, the differential expression was assessed by a linear regression method (Limma package, <http://bioconductor.org/packages/release/bioc/html/limma.html>). Our results showing high degree of overlap between genes differentially expressed in the two data sets indicate that the expression differences observed are independent of the analysis methods.

Because 444 genes were differentially expressed between autism and controls in cortex and only 2 genes were differentially expressed between the two groups in cerebellum (FDR < 0.05), we tested whether this difference was due to the smaller number of cerebellum samples, by relaxing the statistical criteria to FDR < 0.25. We found fewer than 10 differentially expressed genes in cerebellum using the relaxed statistical criteria, supporting the conclusion that genome-wide expression changes in autism were more pronounced in cerebral cortex than in cerebellum.

To account for the fact that the control group of DS1 contained samples from a single female whereas the autism DS1 group included four females, we eliminated from differential expression analysis all probes showing evidence of gender-specific gene expression (*n* = 70). We also applied linear regression of expression values against age and sex, and then assessed differential expression between the autism and control groups using the residual values. We observed a 96% overlap between differentially expressed genes using either the residual values or the raw data, indicating that neither age nor sex were major drivers of expression differences between the autism and control groups.

Differential expression between frontal and temporal cortex was assessed by a paired modified *t*-test (SAM) using the 13 autism and 13 control cases for which RNA samples from both cortex areas passed the quality control criteria. For each of the 510 genes that were differentially expressed in control samples between

frontal and temporal cortex, we compared the variance of autism and control expression values in frontal cortex and temporal cortex. The homogeneity of variance (homoscedasticity) of gene expression was assessed using the Barlett test in R. Fifty one genes showed a significant difference in variance (*P* < 0.05, Barlett test) between autism and control groups both in frontal and temporal cortex, and the Barlett test *P*-values for these genes are listed in Supplementary Data.

**WGCNA.** Unsigned co-expression networks were built using the WGCNA package in R. Probes with evidence of robust expression (9,914; see above) were included in the network. Network construction was performed using the blockwiseModules function in the WGCNA package<sup>29</sup>, which allows the network construction for the entire data set. For each set of genes a pair-wise correlation matrix is computed, and an adjacency matrix is calculated by raising the correlation matrix to a power. The power of 10 was chosen using the scale-free topology criterion<sup>9</sup> and was used for all three networks: the network built using autism samples only, controls samples only or all samples. An advantage of weighted correlation networks is the fact that the results are highly robust with respect to the choice of the power parameter. For each pair of genes, a robust measure of network interconnectedness (topological overlap measure) was calculated based on the adjacency matrix. The topological overlap based dissimilarity was then used as input for average linkage hierarchical clustering. Finally, modules were defined as branches of the resulting clustering tree. To cut the branches, we used the hybrid dynamic tree-cutting because it leads to robustly defined modules<sup>31</sup>. To obtain moderately large and distinct modules, we set the minimum module size to 40 genes and the minimum height for merging modules at 0.1. Each module was summarized by the first principal component of the scaled (standardized) module expression profiles. Thus, the module eigengene explains the maximum amount of variation of the module expression levels. For each module, we defined the module membership measure (also known as module eigengene based connectivity kME) as the correlation between gene expression values and the module eigengene. Genes were assigned to a module if they had a high module membership to the module (kME > 0.7). An advantage of this definition (and the kME measure) is that it allows genes to be part of more than one module. Genes that did not fulfil these criteria for any of the modules are assigned to the grey module. For the cell type marker enrichment analysis we used the markers defined experimentally in refs 32 and 33 which were previously used to annotate human brain network modules<sup>34,35</sup>.

Module visualization: the topological overlap measure was calculated for the top 100 genes in each module ranked by kME. The resulting list of gene pairs was filtered so that both genes in a pair had the highest kME for the module plotted (that is, most module-specific interactions). The resulting top 150 gene pairs were plotted using Visant.

**Gene ontology analyses.** Functional enrichment was assessed using the DAVID database <http://david.abcc.ncifcrf.gov/>. For differentially expressed genes and co-expression modules, the background was set to the total list of genes expressed in the brain in the cortex data set. For genes containing differentially spliced exons, the background was set to the total set of genes showing evidence of alternative splicing in our RNA-seq data. The statistical significance threshold level for all gene ontology enrichment analyses was *P* < 0.05 (Benjamini and Hochberg corrected for multiple comparisons).

**Statistical analyses.** All gene set overlap analyses were performed by assessing the cumulative hypergeometric probability using the phyper function in R. The population size was defined as the total number of probes expressed in both data sets. If the comparison involved different platforms, the comparison was done at gene level.

**Quantitative RT-PCR.** One microgram of total RNA was treated with RNase-free DNase I (Invitrogen/Fermentas) and reverse-transcribed using Invitrogen Superscript II reverse-transcriptase and random hexanucleotide primers (Invitrogen). Real time PCR was performed on an ABI7900 cyclor in 10 µl volume containing iTaq Sybrgreen (Biorad) and primers at a concentration of 0.5 µM each. The results shown in Supplementary Fig. 2b represent at least two independent cDNA synthesis experiments for each gene. *GAPDH* levels were used as an internal control. Statistical significance was assessed by a two-tailed *t*-test assuming unequal variance.

**Semi-quantitative RT-PCR.** Total RNA (600 ng) pooled from autism cases (*n* = 2–3) or controls (*n* = 2–3) was reverse-transcribed as described above. cDNA (50 ng) was subjected to 30 cycles of PCR amplification using the primers described in Supplementary Table 11. PCR products were separated on a 3% agarose gel stained with GelStar (Lonza).

**RNA sequencing and data analysis.** 73-nucleotide reads were generated using an Illumina GAI sequencer according to the manufacturer's protocol. To generate sufficient read coverage for the quantitative analysis of alternative splicing events, reads for ASD and control brain samples were separately pooled and aligned to an existing database of EST and cDNA-derived alternative splicing junctions using



the Basic Local Alignment Tool (BLAT) as described previously<sup>36,37</sup>. Reads were considered properly aligned to a splice junction if at least 71 of the 73 nucleotides matched and at least 5 nucleotides mapped to each of the two exons forming the splice junction. Alternative exon inclusion values ('%inc'), representing the proportion of messenger RNA transcripts with the alternatively spliced exon included, were calculated for each mRNA pool as the ratio of reads aligning to the C1-A or A-C2 junctions against reads aligning against all three possible junctions as previously described<sup>36</sup> (C1-A, A-C2, C1-C2, see Supplementary Fig. 3). Calculated %inc values were considered reliable if at least one of the included junctions as well as the skipped junctions were covered by at least 20 reads. %inc values were compared across samples using Fisher's exact test and the Bonferroni–Hochberg correction to identify differentially spliced exons associated with autism. Differential splicing events were considered significant if they fulfilled both criteria of  $FDR < 0.1$  and %inc difference between autism and controls  $> 15\%$ .

**GWAS set enrichment analysis.** GWAS enrichment analysis was performed as previously described in ref. 38 with the main modification that we generated the null distribution, using permutation of gene labels rather than permutation of case/control labels, because the raw genotyping data was not available for all data sets. This approach has been proposed as an acceptable alternative to phenotype label permutation<sup>38</sup> and has been previously used for set enrichment analyses of GWAS data<sup>39</sup>. For all genes that met the robust expression criteria in our data set, we mapped the SNPs present on the Illumina 550k platform located within the transcript boundaries and an additional 20 kb on the 5' end and 10 kb on the 3' end. Each gene was assigned a GWAS significance value consisting of the lowest  $P$  value of all SNPs mapped to it. A gene set enrichment score (ES) based on the Kolmogorov–Smirnov statistic was calculated as previously described<sup>38</sup> using the  $-\log(P\text{-value})$ . The null distribution was generated by 10,000 random permutations of gene labels in the list of genes/ $P$ -value pairs and an enrichment score ES<sub>p</sub>

was calculated for each permutation. To correct for the gene set size, the enrichment scores were scaled by subtracting the mean and dividing by the standard deviation of ES<sub>p</sub>. The resulting  $z$ -scores were used to calculate the significance  $p$  value.

28. Wintle, R. F. *et al.* (2010). A genotype resource for postmortem brain samples from the Autism Tissue Program. *Autism Res.* **4**, 89–97 (2011).
29. Langfelder, P. & Horvath, S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* **9**, 559 (2008).
31. Langfelder, P., Zhang, B. & Horvath, S. Defining clusters from a hierarchical cluster tree: the Dynamic Tree Cut package for R. *Bioinformatics* **24**, 719–720 (2008).
32. Cahoy, J. D. *et al.* A transcriptome database for astrocytes, neurons, and oligodendrocytes: a new resource for understanding brain development and function. *J. Neurosci.* **28**, 264–278 (2008).
33. Albright, A. V. & Gonzalez-Scarano, F. Microarray analysis of activated mixed glial (microglia) and monocyte-derived macrophage gene expression. *J. Neuroimmunol.* **157**, 27–38 (2004).
34. Oldham, M. C. *et al.* Functional organization of the transcriptome in human brain. *Nature Neurosci.* **11**, 1271–1282 (2008).
35. Miller, J. A., Horvath, S. & Geschwind, D. H. Divergence of human and mouse brain transcriptome highlights Alzheimer disease pathways. *Proc. Natl Acad. Sci. USA* **107**, 12698–12703 (2010).
36. Luco, R. F. *et al.* Regulation of alternative splicing by histone modifications. *Science* **327**, 996–1000 (2010).
37. Pan, Q., Shai, O., Lee, L. J., Frey, B. J. & Blencowe, B. J. Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nature Genet.* **40**, 1413–1415 (2008).
38. Wang, K., Li, M. & Bucan, M. Pathway-based approaches for analysis of genomewide association studies. *Am. J. Hum. Genet.* **81**, 1278–1283 (2007).
39. Zhang, K., Cui, S., Chang, S., Zhang, L. & Wang, J. i-GSEA4GWAS: a web server for identification of pathways/gene sets associated with traits by applying an improved gene set enrichment analysis to genome-wide association study. *Nucleic Acids Res.* **38** (suppl. 2), W90–W95 (2010).

# Detection of prokaryotic mRNA signifies microbial viability and promotes immunity

Leif E. Sander<sup>1</sup>, Michael J. Davis<sup>2\*</sup>, Mark V. Boekschoten<sup>3\*</sup>, Derk Amsen<sup>4</sup>, Christopher C. Dascher<sup>1</sup>, Bernard Ryffel<sup>5</sup>, Joel A. Swanson<sup>2</sup>, Michael Müller<sup>3</sup> & J. Magarian Blander<sup>1</sup>

Live vaccines have long been known to trigger far more vigorous immune responses than their killed counterparts<sup>1–6</sup>. This has been attributed to the ability of live microorganisms to replicate and express specialized virulence factors that facilitate invasion and infection of their hosts<sup>7</sup>. However, protective immunization can often be achieved with a single injection of live, but not dead, attenuated microorganisms stripped of their virulence factors. Pathogen-associated molecular patterns (PAMPs), which are detected by the immune system<sup>8,9</sup>, are present in both live and killed vaccines, indicating that certain poorly characterized aspects of live microorganisms, not incorporated in dead vaccines, are particularly effective at inducing protective immunity. Here we show that the mammalian innate immune system can directly sense microbial viability through detection of a special class of viability-associated PAMPs (vita-PAMPs). We identify prokaryotic messenger RNA as a vita-PAMP present only in viable bacteria, the recognition of which elicits a unique innate response and a robust adaptive antibody response. Notably, the innate response evoked by viability and prokaryotic mRNA was thus far considered to be reserved for pathogenic bacteria, but we show that even non-pathogenic bacteria in sterile tissues can trigger similar responses, provided that they are alive. Thus, the immune system actively gauges the infectious risk by searching PAMPs for signatures of microbial life and thus infectivity. Detection of vita-PAMPs triggers a state of alert not warranted for dead bacteria. Vaccine formulations that incorporate vita-PAMPs could thus combine the superior protection of live vaccines with the safety of dead vaccines.

We hypothesized that the innate immune system might sense the most fundamental characteristic of microbial infectivity, microbial viability itself, and activate a robust immune response regardless of the presence of more specialized factors that regulate microbial virulence<sup>7</sup>. To study the sensing of bacterial viability without the compounding effects of replication or virulence factors, we used thymidine auxotrophs of non-pathogenic *Escherichia coli* K12, strain DH5 $\alpha$  (hereafter called *thyA*<sup>−</sup> *E. coli*). Viable and heat-killed *thyA*<sup>−</sup> *E. coli* similarly activated nuclear factor- $\kappa$ B (NF- $\kappa$ B) and mitogen-activated protein kinase p38 (Supplementary Fig. 1) in murine bone-marrow-derived macrophages and elicited production of similar amounts of interleukin-6 (IL-6) and tumour necrosis factor- $\alpha$  (TNF- $\alpha$ ) (Fig. 1a). In contrast, viable *thyA*<sup>−</sup> *E. coli* induced higher levels of IFN- $\beta$  than heat-killed *thyA*<sup>−</sup> *E. coli* or lipopolysaccharide (LPS) (Fig. 1b), and only viable *thyA*<sup>−</sup> *E. coli* induced IL-1 $\beta$  secretion (Fig. 1c and Supplementary Fig. 2). Pro-IL-1 $\beta$  transcription was equally induced by both viable and heat-killed *thyA*<sup>−</sup> *E. coli* (Fig. 1c), indicating that viable bacteria specifically elicit cleavage of pro-IL-1 $\beta$ . This process is catalysed by caspase-1 in Nod-like receptor (NLR)-containing inflammasome complexes, the assembly of which can be triggered by the activity of bacterial virulence factors<sup>10,11</sup>. Notably, avirulent viable but not heat-killed *thyA*<sup>−</sup> *E. coli* induced inflammasome

activation and pro-caspase-1 cleavage (Fig. 1d). Finally, viable but not heat-killed *thyA*<sup>−</sup> *E. coli* induced caspase-1-dependent inflammatory cell death, termed pyroptosis<sup>10,11</sup>, resulting in the release of lactate dehydrogenase (LDH) (Fig. 1e) and the appearance of 7-amino-actinomycin D (7AAD)<sup>+</sup> annexin-V<sup>−/low</sup> cells (Fig. 1f). Similar responses were observed in peritoneal macrophages and both splenic and bone-marrow-derived dendritic cells (Supplementary Fig. 2b). Killing *thyA*<sup>−</sup> *E. coli* by ultraviolet irradiation, antibiotics, or ethanol also selectively abrogated IL-1 $\beta$  secretion and pyroptosis without affecting IL-6 production (Fig. 1g and Supplementary Fig. 3), indicating that a general determinant associated with bacterial viability is detected.

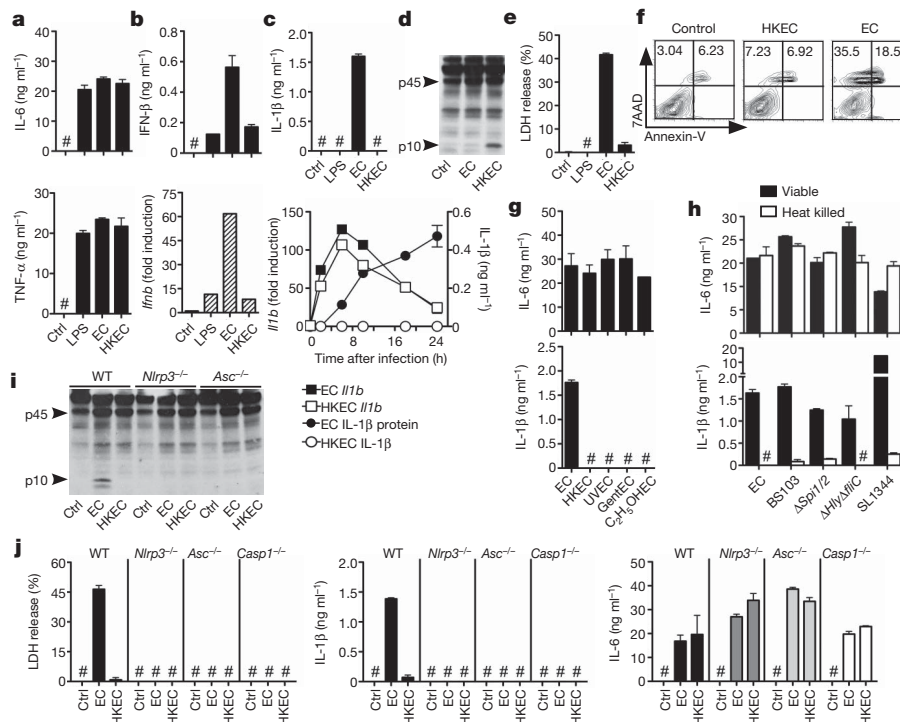
To determine whether pathogenic bacteria can also activate the inflammasome in the absence of virulence factors, we studied attenuated strains of selected pathogens: *Shigella flexneri* virulence plasmid-cured strain BS103<sup>12</sup>, *Salmonella enterica* serovar Typhimurium SL1344  $\Delta$ *Spi1* $\Delta$ *Spi2*, lacking the *Salmonella* pathogenicity islands SPI-1 and SPI-2 (ref. 10), and *Listeria monocytogenes*  $\Delta$ *Hly* $\Delta$ *fliC*, lacking listeriolysin O and flagellin<sup>10</sup>. These mutants induced IL-1 $\beta$  production at levels comparable to those induced by *thyA*<sup>−</sup> *E. coli* (Fig. 1h), but lower and with slower kinetics than their pathogenic counterparts (Supplementary Fig. 4). IL-1 $\beta$  production was abolished when these bacteria were killed, whereas IL-6 production was similar (Fig. 1h). Thus, immune cells detect universal characteristics of viability different from virulence factors.

Caspase-1 activation, pyroptosis and IL-1 $\beta$  production in response to *thyA*<sup>−</sup> *E. coli* were abrogated in macrophages deficient for NLRP3 or for the inflammasome adaptor apoptosis speck protein with caspase recruitment (ASC or PYCARD)<sup>11</sup> (Fig. 1i, j), whereas NLRC4 was dispensable (Supplementary Fig. 5). Pyroptosis and IL-1 $\beta$  production induced by viable *thyA*<sup>−</sup> *E. coli* were abrogated in *Casp1*<sup>−/−</sup> macrophages (Fig. 1j) and suppressed by inhibitors for caspase-1, but not caspase-8 (Supplementary Fig. 6).

Induction of IFN- $\beta$  mRNA and protein by viable *thyA*<sup>−</sup> *E. coli* required the Toll-like receptor (TLR) adaptor TRIF<sup>9</sup> (Fig. 2a, b) and downstream interferon regulatory factor-3 (IRF3)<sup>9</sup> (Supplementary Fig. 7), but not MyD88, the main TLR adaptor<sup>9</sup> (Fig. 2a, b). In contrast, transcription of pro-IL-1 $\beta$  was largely dependent on MyD88. Consequently, *Myd88*<sup>−/−</sup> cells secreted no IL-1 $\beta$  (Fig. 2c, d), whereas pyroptosis and caspase-1 cleavage were intact (Fig. 2e, f). Notably, although TRIF was dispensable for pro-IL-1 $\beta$  transcription (Fig. 2c), *Trif*<sup>−/−</sup> cells failed to secrete IL-1 $\beta$  (Fig. 2d), were protected from pyroptosis (Fig. 2e) and did not activate caspase-1 (Fig. 2f). These findings revealed an unexpected role for TRIF in NLRP3 inflammasome activation in response to viable *thyA*<sup>−</sup> *E. coli*. In contrast, pyroptosis induced by pathogenic *S. enterica* Typhimurium<sup>10</sup> proceeded independently of TRIF (Supplementary Fig. 8). Differential involvement of TRIF, together with differences in magnitude and kinetics of the response (Fig. 1h and Supplementary Fig. 4), indicated that inflammasome activation in

<sup>1</sup>Immunology Institute, Department of Medicine, Mount Sinai School of Medicine, 1425 Madison Avenue, New York, New York 10029, USA. <sup>2</sup>Department of Microbiology and Immunology, University of Michigan, Ann Arbor, Michigan 48109-0620, USA. <sup>3</sup>Nutrition, Metabolism and Genomics Group, Division of Human Nutrition, Wageningen University, 6703 HD Wageningen, The Netherlands. <sup>4</sup>Department of Cell Biology and Histology, Academic Medical Center, University of Amsterdam, 1105 AZ Amsterdam, The Netherlands. <sup>5</sup>Laboratory of Molecular Immunology and Embryology, University of Orleans and Centre National de la Recherche Scientifique, 45071 Orleans, France.

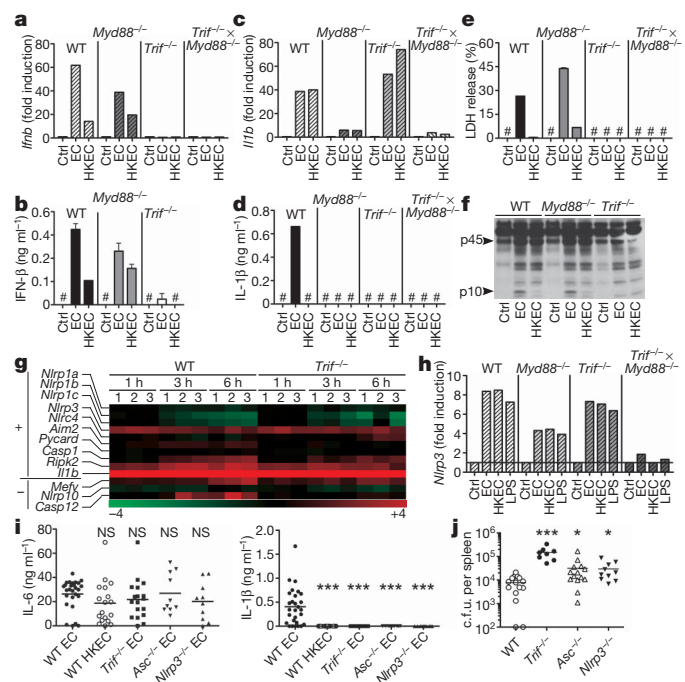
\*These authors contributed equally to this work.



**Figure 1 | Sensing bacterial viability induces IFN-β and activates the NLRP3 inflammasome in the absence of virulence factors.** **a**, **b**, IL-6 and TNF-α (a) and IFN-β protein and mRNA (at 2 h) (b) levels in murine BMs stimulated with medium (ctrl), lipopolysaccharide (LPS), *thyA*<sup>-</sup> *E. coli* (EC) and heat-killed *thyA*<sup>-</sup> *E. coli* (HKEC). Multiplicity of infection = 20. **c**, IL-1β (top), *Il1b* mRNA (bottom, left y axis) and secreted IL-1β (bottom, right y axis) at indicated times is shown. **d**, **i**, Caspase-1 immunoblots at 18 h in wild-type (d) or wild-type (WT), *Nlrp3*<sup>-/-</sup> and *Asc*<sup>-/-</sup> BMs (i). **e**, **f**, Pyroptosis by LDH release (e) and FACS (f) at 18 h is shown. **g**, **h**, IL-6 and IL-1β in response to

response to virulence factors occurs in a manner distinct from that to viability.

Genome-wide transcriptional analysis of wild-type and *Trif*<sup>-/-</sup> macrophages before and after phagocytosis of viable *thyA*<sup>-</sup> *E. coli* showed differential regulation of several clusters of genes (Supplementary Fig. 9) including IFN-regulated genes, as expected<sup>9</sup> (Fig. 2a, b and



*thyA*<sup>-</sup> *E. coli*, viable or killed by different means (g, bone-marrow-derived dendritic cells (BMDCs) or viable or heat-killed *thyA*<sup>-</sup> *E. coli*, attenuated *thyA*<sup>-</sup> *Shigella* (BS103), *Salmonella* ( $\Delta$ *Sp1*/2) and *Listeria* ( $\Delta$ *Hly* $\Delta$ *flc*), or virulent *Salmonella* SL1344 (h). C<sub>2</sub>H<sub>5</sub>OHEC, ethanol-killed *E. coli*; GentEC, gentamicin-killed *E. coli*; UVEEC, UV irradiated *E. coli*. **j**, LDH, IL-1β and IL-6 in BMs of the indicated genotype in response to medium (ctrl), viable *thyA*<sup>-</sup> *E. coli* (EC) and heat-killed *thyA*<sup>-</sup> *E. coli* (HKEC). All responses are by murine BMs and measured at 24 h unless indicated otherwise. Hash symbol indicates not detected. Data represent ≥5 experiments. All bars represent mean ± s.e.m.

Supplementary Fig. 10a), whereas most of the Rel/NF-κB target genes were comparable (Supplementary Fig. 10b). *Nlrp3* expression was induced independently of TRIF (Fig. 2g, h), and negative regulators of inflammasome activity, such as those encoded by Mediterranean fever (*Mefv*), *Nlrp10* and *Casp12* genes, were also unchanged or expressed at higher levels in wild-type macrophages (Fig. 2g), possibly due to negative feedback. Thus, the role of TRIF in inflammasome activation upon phagocytosis of viable *thyA*<sup>-</sup> *E. coli* is not explained by transcriptional control of inflammasome components (so called priming<sup>11</sup>). Furthermore, ATP and reactive oxygen species (ROS)<sup>11,13</sup>, known activators of the NLRP3 inflammasome, were not involved, as deficiency for P<sub>2</sub>X<sub>7</sub>R, which is required for ATP-mediated NLRP3 activation, did not affect pyroptosis or IL-1β production (Supplementary Fig. 11a, b), and ROS accumulated equally in response to viable and heat-killed *thyA*<sup>-</sup> *E. coli* independently of TRIF (Supplementary Fig. 11c).

**Figure 2 | The TLR signalling adaptor TRIF controls 'viability-induced' responses.** **a–e**, *Il1b* transcription at 2 h (a), IFN-β secretion at 24 h (b), *Il1b* transcription at 2 h (c), IL-1β secretion (d) and LDH release (e) at 24 h after phagocytosis of viable (EC) or heat-killed (HKEC) *thyA*<sup>-</sup> *E. coli*. **f**, Caspase-1 immunoblot at 18 h. Data in **a–f** are from murine BMs and represent ≥5 experiments. **g**, Gene microarray analysis of wild-type and *Trif*<sup>-/-</sup> BMs treated with viable *thyA*<sup>-</sup> *E. coli* for 1, 3 or 6 h (three biological replicates, numbered 1–3). A heat map of positive regulators/essential components (+) and negative regulators (-) of inflammasomes is shown. **h**, *Nlrp3* transcription at 1 h in BMs. **i**, **j**, Serum levels of IL-6 and IL-1β 6 h after injection of  $1 \times 10^9$  viable or  $5 \times 10^9$  heat-killed *thyA*<sup>-</sup> *E. coli* (i), and splenic bacterial burdens 72 h after injection of  $1 \times 10^8$  non-auxotroph *E. coli* (j) into wild-type, *Trif*<sup>-/-</sup>, *Asc*<sup>-/-</sup> and *Nlrp3*<sup>-/-</sup> mice are shown. Each symbol represents one mouse. \*,  $P \leq 0.05$ ; \*\*,  $P \leq 0.01$ ; \*\*\*,  $P \leq 0.001$ . NS, not statistically significant. Hash symbol indicates not detected. All bars represent mean ± s.e.m.



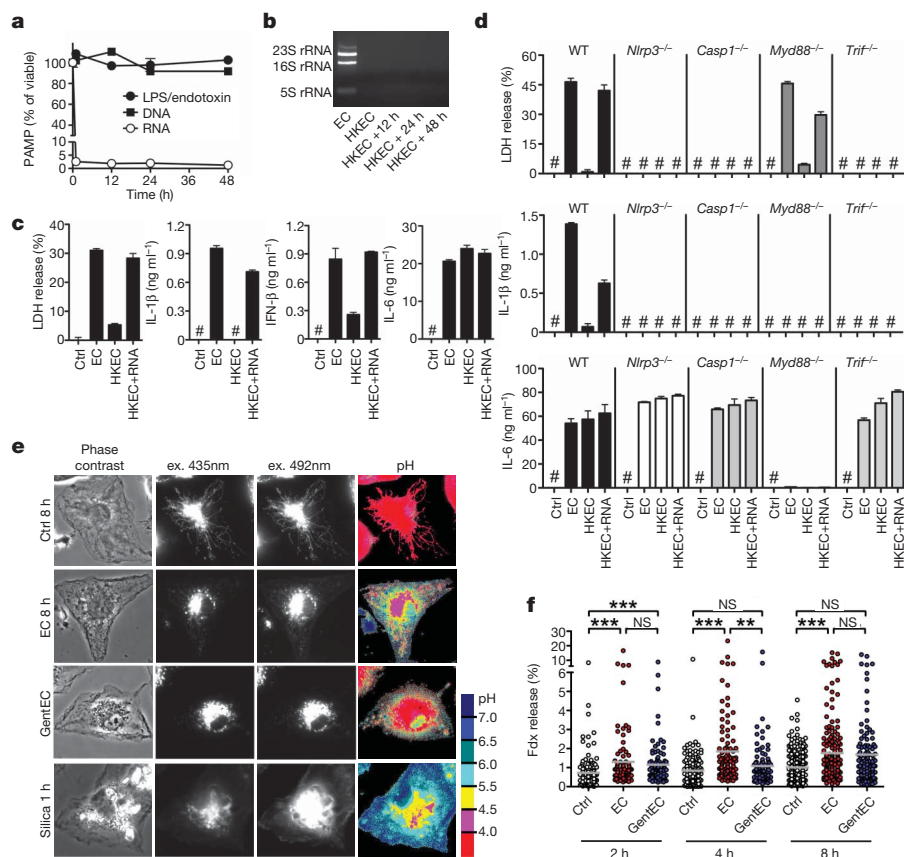
Injection of viable and heat-killed *thyA*<sup>-</sup> *E. coli* into mice induced similarly high serum levels of IL-6 (Fig. 2i). In contrast, circulating IL-1 $\beta$  was detected only in mice infected with viable bacteria (Fig. 2i), whereas IFN- $\beta$  levels were undetectable in all groups (data not shown). Confirming our results *in vitro*, production of IL-1 $\beta$  (but not IL-6) *in vivo* also required TRIF, ASC and NLRP3 (Fig. 2i). Injection of non-pathogenic *S. enterica* Typhimurium induced serum IL-1 $\beta$  levels comparable to those elicited by *thyA*<sup>-</sup> *E. coli*, which similarly depended on TRIF (Supplementary Fig. 12). Although pathogenic *S. enterica* Typhimurium elicited higher levels of serum IL-1 $\beta$  than non-pathogenic *Salmonella*, this response was also severely reduced in *Trif*<sup>-/-</sup> mice, suggesting a previously unappreciated role for TRIF in *Salmonella* infection (Supplementary Fig. 12). Importantly, deficiency in TRIF, ASC and NLRP3 impaired bacterial clearance during systemic infection with replication-sufficient non-pathogenic *E. coli* (Fig. 2j). This failure was more dramatic in *Trif*<sup>-/-</sup> than in *Asc*<sup>-/-</sup> or *Nlrp3*<sup>-/-</sup> mice, possibly due to the central upstream role of TRIF in inflammasome activation and IFN- $\beta$  production.

The ability to sense microbial viability through pathways downstream of pattern recognition receptors indicates the existence of vita-PAMPs; that is, PAMPs associated with viable but not dead bacteria. In contrast to LPS and genomic DNA, which remained constant after killing *thyA*<sup>-</sup> *E. coli* with heat, total bacterial RNA was rapidly lost (Fig. 3a, b and Supplementary Fig. 13). Total RNA content was also lost with antibiotic treatment, and little ribosomal RNA (rRNA) remained after killing with ultraviolet irradiation and ethanol (Supplementary Fig. 14). Only fixation with paraformaldehyde (PFA) efficiently killed the bacteria (not shown) while preserving total RNA content (Supplementary Fig. 15a). Remarkably, unlike bacteria killed by other means, PFA-killed bacteria induced pyroptosis and IL-1 $\beta$  production to levels similar to those induced by viable bacteria (Supplementary Fig. 15b). Thus, the presence or absence of RNA correlated with the ability to activate pathways involved in sensing viability.

These results indicate that prokaryotic RNA represents a labile PAMP closely associated with bacterial viability that might signify

microbial life to the immune system. Indeed, addition of purified total bacterial RNA fully restored the ability of heat-killed *thyA*<sup>-</sup> *E. coli* to induce pyroptosis, IL-1 $\beta$  and IFN- $\beta$  production (Fig. 3c). These responses were dependent on TRIF, NLRP3 and caspase-1, just as those responses elicited by viable bacteria (Fig. 3d compared to Figs 1j and 2a–f). The NLRP3 inflammasome mediates recognition of viral RNA during influenza A infection<sup>14</sup>. Together with our results and those of others<sup>15</sup>, this suggests a more general role for NLRP3 in responses to RNAs of microbial origin. RNA can activate the NLRP3 inflammasome when delivered into the cytosol (where NLRP3 is found) with transfection reagents<sup>15</sup>. In contrast, inflammasome activation by the combination of total bacterial RNA and dead *thyA*<sup>-</sup> *E. coli* did not require RNA transfection (Fig. 3c, d). Administration of total *E. coli* RNA alone or in combination with LPS (to mimic an *E. coli*-derived PAMP plus RNA) had little effect on NLRP3 inflammasome activation unless the RNA was delivered to the cytosol using Lipofectamine (Supplementary Fig. 16) or in combination with ATP, as reported previously<sup>15</sup>. Thus, phagocytosis of viable bacteria is a natural context of bacterial-RNA-mediated NLRP3 inflammasome activation.

These findings raised the question as to how vita-PAMPs in phagolysosomes gain access to cytosolic receptors such as NLRP3 in the absence of invasion, auxiliary secretion systems or pore-forming toxins. To address this question, we exploited the pH-sensitive excitation spectrum of fluorescein: the acidic pH in phagolysosomes quenches fluorescence whereas release into the pH-neutral cytosol allows a regain in fluorescence<sup>16</sup>. Phagocytosis of avirulent *thyA*<sup>-</sup> *E. coli* in the presence of fluorescein-conjugated dextran (Fdx) consistently induced low-level release of Fdx into the cytosol of macrophages (Fig. 3e, f and Supplementary Fig. 17). This indicates that phagosomes carrying *E. coli* exhibit intrinsic leakiness, a property previously described for particles such as beads and crystals that induce phagolysosomal destabilization<sup>16,17</sup>. Interestingly, killed *E. coli* also induced Fdx release, although to a slightly lower extent than viable *E. coli* (Fig. 3e, f), demonstrating that phagosomal leakage occurs independently of bacterial viability. Therefore, RNA from viable



**Figure 3 | Bacterial RNA is a vita-PAMP that accesses cytosolic receptors during phagocytosis and in the absence of virulence factors.** **a**, LPS/endotoxin, genomic DNA and total RNA in *thyA*<sup>-</sup> *E. coli* before and at indicated times after heat killing. **b**, Agarose gel electrophoresis of *thyA*<sup>-</sup> *E. coli* total RNA before and after heat killing at 60 °C for 60 min followed by 4 °C incubation for the indicated times. **c**, **d**, LDH, IL-1 $\beta$ , IFN- $\beta$  and IL-6 at 24 h in response to viable *thyA*<sup>-</sup> *E. coli* (EC), heat-killed *thyA*<sup>-</sup> *E. coli* (HKEC), or heat-killed *thyA*<sup>-</sup> *E. coli* with 10  $\mu$ g ml<sup>-1</sup> total RNA (HKEC+RNA). Hash symbol in **c** and **d** indicates not detected. Data in **a**–**d** are from murine BMs and represent  $\geq 5$  experiments. **e**, Representative ratiometric epifluorescence imaging of murine BMs at 8 h with Fdx alone (ctrl 8 h), Fdx and viable *thyA*<sup>-</sup> *E. coli* (EC 8 h) or gentamicin-killed *thyA*<sup>-</sup> *E. coli* (GentEC). Colour code indicates pH scale. Positive control is ground silica (silica 1 h). **f**, Quantification of cytosolic Fdx expressed as percentage of total Fdx per cell. Each dot represents the percentage of released Fdx per individual cell. Grey bars represent mean Fdx release. \**P* < 0.05; \*\**P* < 0.01; \*\*\**P* < 0.001. All bars represent mean  $\pm$  s.e.m.

bacteria could gain access to cytosolic receptors via intrinsic phagosomal leakage. These results may also explain the reported ability of phagosome-degraded mutants of *Listeria monocytogenes* or *Staphylococcus aureus* to induce a transcriptional response dependent on cytosolic NLRs<sup>18,19</sup>.

Digestion of total RNA from *E. coli* with exonuclease RNase I and double-stranded RNA (dsRNA)-specific endonuclease RNase III abrogated LDH and IL-1 $\beta$  release, whereas DNase treatment had no effect (Fig. 4a). Of the *E. coli* RNA species, mRNA most potently induced pyroptosis as well as production of IL-1 $\beta$  and IFN- $\beta$ . Small RNA (sRNA), or the most abundant RNA, ribosomal RNA (rRNA), had little or no detectable effects (Fig. 4b and Supplementary Fig. 18). *Escherichia coli* rRNA undergoes extensive modifications not found in mRNA<sup>20</sup>, which may underlie the differential activity of these RNA species. The relative amount of mRNA was <1% of the total RNA and accordingly, mRNA was approximately 100-fold more effective than total RNA (Figs 3c and 4a, b and Supplementary Fig. 18).

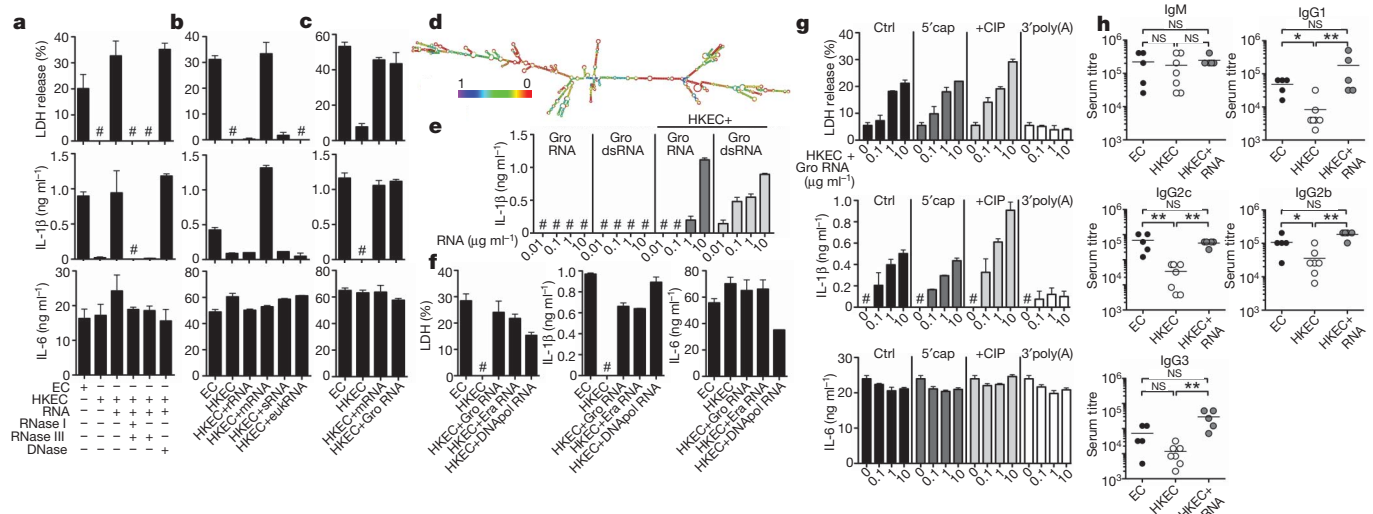
*In-vitro*-transcribed single-stranded mRNA of the *E. coli* Gro operon (Supplementary Fig. 19a, b), which is strongly expressed upon phagocytosis of bacteria<sup>21</sup>, induced caspase-1 cleavage and subsequent pyroptosis and IL-1 $\beta$  production when phagocytosed together with heat-killed *thyA*<sup>-</sup> *E. coli* (Fig. 4c, e and Supplementary Fig. 19c–e). The single-stranded Gro mRNA sequence had a predicted secondary structure with regions of high probability for base pairing (Fig. 4d), consistent with susceptibility of the stimulatory activity to RNase III treatment (Fig. 4a). Indeed, fully dsGro mRNA (Supplementary Fig. 19b) induced responses similar to single-stranded Gro mRNA of the appropriate length (Fig. 4e and Supplementary Fig. 19d). Other transcripts also induced such responses, showing that the immunostimulatory property is independent of RNA sequence (Fig. 4f).

Notably, eukaryotic RNA was unable to elicit the responses induced by *E. coli* mRNA (Fig. 4b). Unlike eukaryotic mRNA, triphosphate moieties at the 5' end of bacterial mRNAs are not capped with 7-methyl-guanosine (7m<sup>7</sup>G)<sup>22</sup>, and might betray the prokaryotic origin of these transcripts<sup>23</sup>. However, neither treatment with calf

intestinal phosphatase (CIP) nor capping affected the activity of Gro mRNA during phagocytosis of heat-killed *thyA*<sup>-</sup> *E. coli* (Fig. 4g). The stimulatory activity of purified *E. coli* total RNA or mRNA was also unaltered by CIP treatment (Supplementary Fig. 20a, b), arguing against a role for the RNA helicase retinoic acid inducible gene-I (RIG-I), which can induce interferon and IL-1 $\beta$  production but requires 5'-triphosphates for activation (Supplementary Fig. 20c)<sup>23</sup>. Moreover, TRIF and NLRP3 are dispensable for RIG-I function but are required for the stimulatory activity of bacterial RNA (Figs 2a, b and 3d). Interestingly, RNA can induce RIG-I-dependent IFN- $\beta$  during infection with an invasive intracellular bacterium<sup>24</sup>, indicating that the nature of microbial pathogenesis and the cellular context in which bacterial RNA is recognized may determine the choice of innate sensors engaged. In contrast to 5'-triphosphate removal, adding polyadenyl groups to the 3' end of Gro mRNA or purified *E. coli* mRNA abrogated IL-1 $\beta$  secretion and pyroptosis (Fig. 4g and Supplementary Fig. 21). Thus, absence of 3'-polyadenylation<sup>22</sup> may allow specific detection of prokaryotic mRNA during infection. Additional features may distinguish self from microbial RNAs such as internal naturally occurring nucleoside modifications in eukaryotic RNA<sup>25–27</sup>.

To test the impact of vita-PAMPs on adaptive immunity, we immunized mice with either viable or dead *thyA*<sup>-</sup> *E. coli*, or a combination of dead *thyA*<sup>-</sup> *E. coli* and purified total bacterial RNA (Supplementary Fig. 22). Whereas all three vaccines induced similar polyclonal anti-*E. coli* IgM responses, production of class-switched IgG subclasses was strongly enhanced in response to vaccination with viable compared to killed *E. coli* (Fig. 4h). Adding total bacterial RNA to killed *thyA*<sup>-</sup> *E. coli* elevated IgG1, IgG2c, IgG2b and IgG3 antibody titres to or above the levels in mice immunized with viable *thyA*<sup>-</sup> *E. coli*. Thus, innate detection of bacterial viability leads to robust activation of a humoral adaptive response. These findings indicate that bacterial RNA can augment killed vaccines to perform as well as live ones.

Our findings reveal an inherent ability of the immune system to distinguish viable from dead microorganisms. The presence of live bacteria in sterile tissues, regardless of whether these (still) express



**Figure 4 | Bacterial mRNA constitutes an active vita-PAMP.** **a–c, e–g,** LDH, IL-1 $\beta$  and IL-6 at 24 h. **a,** Total *thyA*<sup>-</sup> *E. coli* RNA treated with RNase I and RNase III, RNase III alone, or DNase before stimulation of BMDCs. **b,** BMDCs treated with viable or heat-killed *thyA*<sup>-</sup> *E. coli*, or heat-killed *thyA*<sup>-</sup> *E. coli* with 0.1  $\mu\text{g ml}^{-1}$  of different bacterial RNA (ribosomal RNA (rRNA), mRNA, small RNA (sRNA) or eukaryotic RNA (eukRNA)). **c,** BMDC responses. Gro RNA indicates *in-vitro*-transcribed *E. coli* Gro operon RNA. **d,** Predicted secondary structure of Gro RNA. The colour code indicates base pairing probability. **e,** BMDCs treated with *in-vitro*-transcribed Gro RNA or Gro dsRNA alone or with heat-killed *thyA*<sup>-</sup> *E. coli*. **f,** BMDC responses. Era RNA and DNAPol RNA indicate *in-vitro*-transcribed *E. coli* Era GTPase and DNA polymerase III RNA, respectively. **g,** BMDCs treated with different doses of unmodified (ctrl, control) or modified Gro RNA with heat-killed *thyA*<sup>-</sup> *E. coli* (5' cap, 5' m<sup>7</sup>G capping; CIP, calf intestinal phosphatase; 3' poly(A), 3'-polyadenylation). For **a–g**, the hash symbol indicates not detected; all RNA at 10  $\mu\text{g ml}^{-1}$  except as noted; data represent  $\geq 5$  experiments. **h,** Mice vaccinated and boosted twice with viable *thyA*<sup>-</sup> *E. coli* (EC), heat-killed *thyA*<sup>-</sup> *E. coli* (HKEC) or heat-killed *thyA*<sup>-</sup> *E. coli* with 30  $\mu\text{g}$  total purified bacterial RNA (HKEC+RNA) (vaccination regimen is given in Supplementary Fig. 22). Class-specific anti-*E. coli* antibody serum titres at 25 days are shown. \*,  $P \leq 0.05$ ; \*\*,  $P \leq 0.01$ ; \*\*\*,  $P \leq 0.001$ . All bars represent mean  $\pm$  s.e.m.

virulence factors, poses an acute threat that must be dealt with by an aggressive immune response. Dead bacteria, on the other hand, would signify a successful immune response that can now subside. Detection of vita-PAMPs within sterile tissues signifies microbial viability. Other vita-PAMPs may exist in the form of second messengers like cyclic diadenosine or di-guanosine monophosphates<sup>7,28</sup> or quorum-sensing molecules<sup>7</sup>. The extent to which vita-PAMPs contribute to the host response during natural infection with pathogenic bacteria, relative to other stimuli such as the activity of virulence factors, is an important issue that requires further investigation. Given that bacteria tightly regulate their virulence via multiple mechanisms in response to different environmental signals and inside a host organism during infection<sup>29,30</sup>, detection of invariant vita-PAMPs essential to bacterial survival may be a non-redundant fail-safe strategy for host protection.

## METHODS SUMMARY

Cells were infected with *E. coli* DH5 $\alpha$  *thyA*<sup>-</sup> at a multiplicity of infection of 20 for 24 h unless stated otherwise. Supernatants were assayed for cytokines by ELISA. Genome-wide transcriptional analysis of murine bone-marrow-derived macrophages (BMMs) at 0, 1, 3 and 6 h after infection was carried out on Affymetrix GeneChip Mouse Gene 1.1 ST 24-array plates. Phagosomal leakage in BMMs was detected by measuring Fdx release using a modified method previously described<sup>16</sup>. In brief, BMMs were treated with *thyA*<sup>-</sup> *E. coli* in the presence of 0.167 mg ml<sup>-1</sup> Fdx and imaged with excitation at 440 nm (pH insensitive) and 485 nm (pH sensitive). Fluorescence intensity ratios at 485 nm/440 nm were converted into pH maps and the percentage of Fdx release calculated (total intensity of pixels containing released Fdx/total Fdx intensity). Bacterial RNA was extracted from *E. coli* using the e.z.n.a RNA kit (Omega) and *in vitro* transcription of bacterial genes carried out using the MEGAscript kit (Ambion) followed by DNase digestion and RNA purification using the MEGAclear kit (Ambion). RNA polyadenylation was performed with the poly(A)-tailing kit (Ambion). Vaccinations were performed as a prime-boost regimen (see Methods). C57BL/6j and *P2rx7*<sup>-/-</sup> mice were purchased from the Jackson Laboratory. *Myd88*<sup>-/-</sup> and *Trif*<sup>-/-</sup> mice were provided by S. Akira, *Trif*<sup>-/-</sup>  $\times$  *Myd88*<sup>-/-</sup> by R. Medzhitov, *Nlrp3*<sup>-/-</sup>, *Asc*<sup>-/-</sup> and *Nlr4*<sup>-/-</sup> by Millenium, and *Casp1*<sup>-/-</sup> by R. Flavell. Animal care and experimentation were performed in accordance with approved MSSM Institutional Animal Care and Use Committee protocols.

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

Received 30 April 2010; accepted 24 March 2011.

Published online 22 May 2011.

1. Brockstedt, D. G. *et al.* Killed but metabolically active microbes: a new vaccine paradigm for eliciting effector T-cell responses and protective immunity. *Nature Med.* **11**, 853–860 (2005).
2. Cheers, C. & Zhan, Y. How do macrophages distinguish the living from the dead? *Trends Microbiol.* **4**, 453–455 (1996).
3. Detmer, A. & Glenting, J. Live bacterial vaccines—a review and identification of potential hazards. *Microb. Cell Fact.* **5**, 23 (2006).
4. Kawamura, I. *et al.* Antigen provoking gamma interferon production in response to *Mycobacterium bovis* BCG and functional difference in T-cell responses to this antigen between viable and killed BCG-immunized mice. *Infect. Immun.* **62**, 4396–4403 (1994).
5. Lauvau, G. *et al.* Priming of memory but not effector CD8 T cells by a killed bacterial vaccine. *Science* **294**, 1735–1739 (2001).
6. von Koenig, C. H., Finger, H. & Hof, H. Failure of killed *Listeria monocytogenes* vaccine to produce protective immunity. *Nature* **297**, 233–234 (1982).
7. Vance, R. E., Isberg, R. R. & Portnoy, D. A. Patterns of pathogenesis: discrimination of pathogenic and nonpathogenic microbes by the innate immune system. *Cell Host Microbe* **6**, 10–21 (2009).
8. Medzhitov, R. Approaching the asymptote: 20 years later. *Immunity* **30**, 766–775 (2009).
9. Takeuchi, O. & Akira, S. Pattern recognition receptors and inflammation. *Cell* **140**, 805–820 (2010).
10. Mariathasan, S. & Monack, D. M. Inflammasome adaptors and sensors: intracellular regulators of infection and inflammation. *Nature Rev. Immunol.* **7**, 31–40 (2007).
11. Schroder, K. & Tschopp, J. The inflammasomes. *Cell* **140**, 821–832 (2010).

12. Wing, H. J., Yan, A. W., Goldman, S. R. & Goldberg, M. B. Regulation of IcsP, the outer membrane protease of the *Shigella* actin tail assembly protein IcsA, by virulence plasmid regulators VirF and VirB. *J. Bacteriol.* **186**, 699–705 (2004).
13. Zhou, R., Yazdi, A. S., Menu, P. & Tschopp, J. A role for mitochondria in NLRP3 inflammasome activation. *Nature* **469**, 221–225 (2011).
14. Pang, I. K. & Iwasaki, A. Inflammasomes as mediators of immunity against influenza virus. *Trends Immunol.* **32**, 34–41 (2011).
15. Kanneganti, T. D. *et al.* Bacterial RNA and small antiviral compounds activate caspase-1 through cryopyrin/Nalp3. *Nature* **440**, 233–236 (2006).
16. Davis, M. J. & Swanson, J. A. Technical advance: Caspase-1 activation and IL-1 $\beta$  release correlate with the degree of lysosome damage, as illustrated by a novel imaging method to quantify phagolysosome damage. *J. Leukoc. Biol.* **88**, 813–822 (2010).
17. Hornung, V. *et al.* Silica crystals and aluminum salts activate the NALP3 inflammasome through phagosomal destabilization. *Nature Immunol.* **9**, 847–856 (2008).
18. Herskovits, A. A., Auerbuch, V. & Portnoy, D. A. Bacterial ligands generated in a phagosome are targets of the cytosolic innate immune system. *PLoS Pathog.* **3**, e51 (2007).
19. Shimada, T. *et al.* *Staphylococcus aureus* evades lysozyme-based peptidoglycan digestion that links phagocytosis, inflammasome activation, and IL-1 $\beta$  secretion. *Cell Host Microbe* **7**, 38–49 (2010).
20. Piekna-Przybylska, D., Decatur, W. A. & Fournier, M. J. The 3D rRNA modification maps database: with interactive tools for ribosome analysis. *Nucleic Acids Res.* **36**, D178–D183 (2008).
21. Buchmeier, N. A. & Heffron, F. Induction of *Salmonella* stress proteins upon infection of macrophages. *Science* **248**, 730–732 (1990).
22. Belasco, J. G. All things must pass: contrasts and commonalities in eukaryotic and bacterial mRNA decay. *Nature Rev. Mol. Cell Biol.* **11**, 467–478 (2010).
23. Rehwinkel, J. & Reis e Sousa, C. RIGorous detection: exposing virus through RNA sensing. *Science* **327**, 284–286 (2010).
24. Monroe, K. M., McWhirter, S. M. & Vance, R. E. Identification of host cytosolic sensors and bacterial factors regulating the type I interferon response to *Legionella pneumophila*. *PLoS Pathog.* **5**, e1000665 (2009).
25. Nallagatla, S. R., Toroney, R. & Bevilacqua, P. C. A brilliant disguise for self RNA: 5'-end and internal modifications of primary transcripts suppress elements of innate immunity. *RNA Biol.* **5**, 140–144 (2008).
26. Anderson, B. R. *et al.* Incorporation of pseudouridine into mRNA enhances translation by diminishing PKR activation. *Nucleic Acids Res.* **38**, 5884–5892 (2010).
27. Kariko, K., Buckstein, M., Ni, H. & Weissman, D. Suppression of RNA recognition by Toll-like receptors: the impact of nucleoside modification and the evolutionary origin of RNA. *Immunity* **23**, 165–175 (2005).
28. Woodward, J. J., Iavarone, A. T. & Portnoy, D. A. c-di-AMP secreted by intracellular *Listeria monocytogenes* activates a host type I interferon response. *Science* **328**, 1703–1705 (2010).
29. Gripenland, J. *et al.* RNAs: regulators of bacterial virulence. *Nature Rev. Microbiol.* **8**, 857–866 (2010).
30. Raskin, D. M., Seshadri, R., Pukatzki, S. U. & Mekalanos, J. J. Bacterial genomics and pathogen evolution. *Cell* **124**, 703–714 (2006).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We are grateful to R. Medzhitov and J. C. Kagan for critical reading of the manuscript; C. B. Lopez for *Ir3*<sup>-/-</sup> mice; D. M. Monack for *Salmonella*  $\Delta$ *Spi1*  $\Delta$ *Spi2*; M. B. Goldberg for *Shigella* BS103; and D. A. Portnoy for *Listeria*  $\Delta$ *Hly*  $\Delta$ *flc*. We thank M. Riveccio, I. Brodsky, M. Blander, S. J. Blander, J. Sander and Blander laboratory members for insightful discussions, help and support. L.E.S. was supported by Deutsche Forschungsgemeinschaft grant SA-1940/1-1, D.A. by fellowships from the Academic Medical Center and the Landsteiner Foundation for Blood Research, and M.V.B. and M.M. by the Netherlands Nutrigenomics Centre. This work was supported by NIH grant AI080959A and the Kinship Foundation Searle Scholar award to J.M.B.

**Author Contributions** L.E.S. and J.M.B. designed experiments and directed the study. L.E.S. performed all experiments. M.J.D. and L.E.S. performed experiments measuring lysosomal leakage. J.A.S. helped with the design and analysis of the lysosomal leakage experiments. M.V.B. performed gene microarray analysis. M.V.B. and M.M. analysed the gene microarray data and helped with data interpretation. D.A. and J.M.B. performed experiments during the development phase of the project, and C.C.D. helped with the design of RNA-related experiments. B.R. provided bone marrow progenitor cells from *Nlrp3*<sup>-/-</sup>, *Asc*<sup>-/-</sup> and *Casp1*<sup>-/-</sup> mice. L.E.S., D.A. and J.M.B. wrote the manuscript. J.M.B. conceived of the study.

**Author Information** Affymetrix Microarray data have been deposited with the NCBI Gene Expression Omnibus (<http://www.ncbi.nlm.nih.gov/geo/>) under accession number GSE27960. Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at [www.nature.com/nature](http://www.nature.com/nature). Correspondence and requests for materials should be addressed to J.M.B. ([julie.blander@mssm.edu](mailto:julie.blander@mssm.edu)).



## METHODS

**Cells.** Bone-marrow-derived dendritic cell (BMDC) cultures were grown as previously described<sup>31</sup> in RPMI 1640 supplemented with granulocyte-macrophage colony-stimulating factor (GM-CSF) and 5% fetal bovine serum (FBS), plus 100 µg ml<sup>-1</sup> penicillin, 100 µg ml<sup>-1</sup> streptomycin, 2 mM L-glutamine, 10 mM HEPES, 1 nM sodium pyruvate, 1% MEM non-essential amino acids, and 2.5 µM β-mercaptoethanol (all Sigma). Semi-adherent cells were harvested on ice on day 5 and re-plated immediately in fresh RPMI 1640 medium containing 10% FBS at 5 × 10<sup>5</sup> cells per well in 24-well tissue-culture-treated plates. Stimuli were added immediately after re-plating in the same medium and the cells were centrifuged for 2 min at 2,000 r.p.m. Murine macrophages were derived from the bone marrow (BMMs) of C57BL/6J, *Myd88*<sup>-/-</sup>, *Trif*<sup>-/-</sup>, *Trif*<sup>-/-</sup> × *Myd88*<sup>-/-</sup>, *Nlrp3*<sup>-/-</sup>, *Asc*<sup>-/-</sup> or *Casp1*<sup>-/-</sup> mice, as described previously<sup>32</sup>, in RPMI 1640 supplemented with M-CSF and 10% FBS, plus 100 µg ml<sup>-1</sup> penicillin and 100 µg ml<sup>-1</sup> streptomycin, 10 mM HEPES and 1 nM sodium pyruvate (all Sigma). For some experiments macrophages were derived from the bone marrow of *Irf3*<sup>-/-</sup> or *P2rx7*<sup>-/-</sup> mice. Peritoneal macrophages were harvested 72 h after intraperitoneal injection of 1 ml thioglycollate (BD Bioscience), grown overnight in RPMI 1640 medium supplemented with 10% FBS and 100 µg ml<sup>-1</sup> penicillin and 100 µg ml<sup>-1</sup> streptomycin, hereafter referred to as 'complete medium'. Mouse embryonic fibroblasts (MEFs) deficient for RIG-I (*RIG-I*<sup>-/-</sup>) were provided by A. Ting with permission from S. Akira, and grown in DMEM medium containing 10% FBS and 100 µg ml<sup>-1</sup> penicillin, 100 µg ml<sup>-1</sup> streptomycin.

**Mice.** C57BL/6J and *P2rx7*<sup>-/-</sup> mice were purchased from Jackson Laboratories. *Myd88*<sup>-/-</sup> and *Trif*<sup>-/-</sup> mice were originally provided by S. Akira; *Myd88*<sup>-/-</sup> and *Trif*<sup>-/-</sup> mice were interbred to homozygosity to generate *Trif*<sup>-/-</sup> × *Myd88*<sup>-/-</sup> mice, and were provided by R. Medzhitov. *Nlrp3*<sup>-/-</sup>, *Asc*<sup>-/-</sup> or *Casp1*<sup>-/-</sup> bone marrow was provided by B. Ryffel and mice for *in vivo* studies were acquired from R. Flavell (through Millenium) and have been described previously<sup>33,34</sup>. *Irf3*<sup>-/-</sup> mice were provided by C. B. Lopez and were previously described<sup>35</sup>. We used 8–10-week-old animals for all experiments. All experiments were approved by the institutional ethics committee and carried out in agreement with the 'Guide for the Care and Use of Laboratory Animals' (NIH publication 86-23, revised 1985).

**Bacteria.** *Escherichia coli* K12, strain DH5α were purchased from Invitrogen. Naturally occurring thymidine auxotrophs (*thyA*<sup>-</sup>) were selected on Luria-Bertani (LB) agar plates containing 50 µg ml<sup>-1</sup> trimethoprim and 500 µg ml<sup>-1</sup> thymidine (both Sigma). Auxotrophy was confirmed by inoculation and overnight culture of single colonies in LB medium. *thyA*<sup>-</sup> *E. coli* grew only in the presence of thymidine and were resistant to trimethoprim. For phagocytosis experiments, *thyA*<sup>-</sup> *E. coli* were grown to mid-log phase, washed three times in phosphate buffered saline (PBS) to remove thymidine and LB salts before addition to cells. For heat killing, *thyA*<sup>-</sup> *E. coli* were grown to log phase, washed and re-suspended in PBS at an optical density at 600 nm (OD<sub>600</sub>) of 0.6, and subsequently incubated at 60 °C for 60 min. *thyA*<sup>-</sup> heat-killed *E. coli* were stored up to 18 h at 4 °C or used immediately after cooling. Efficient killing was confirmed by overnight plating on thymidine/trimethoprim-supplemented LB-agar plates. For gentamicin killing, *thyA*<sup>-</sup> *E. coli* were grown to mid-log phase, washed and re-suspended in LB medium containing thymidine, trimethoprim and 50 µg ml<sup>-1</sup> gentamicin sulphate and incubated in a shaking incubator at 37 °C overnight. Ethanol killing was carried out by re-suspending log phase *thyA*<sup>-</sup> *E. coli* in 70% ethanol for 10 min, followed by extensive washing in PBS. For ultraviolet killing, log phase *thyA*<sup>-</sup> *E. coli* were re-suspended in PBS at an OD<sub>600</sub> of 0.6, ultraviolet-irradiated with 1,000 mJ cm<sup>-2</sup> in a Petri dish followed by washing with PBS. Paraformaldehyde (PFA) fixation was performed by re-suspending log-phase *thyA*<sup>-</sup> *E. coli* in 4% PFA in PBS for 10 min followed by extensive washing and re-suspension in PBS. *Shigella flexneri* virulence plasmid-cured strain BS103 was provided by M. B. Goldberg<sup>12,36</sup>. *thyA*<sup>-</sup> *S. flexneri* were selected similarly to *thyA*<sup>-</sup> *E. coli*. D. M. Monack provided *Salmonella enterica* serovar Typhimurium, strain SL1344 Δ*Spi1*Δ*Spi2*, lacking the *Salmonella* pathogenicity island SPI-1 and SPI-2 type-III secretion systems<sup>37</sup>. SL1344 Δ*Spi1*Δ*Spi2* was grown in LB medium containing 25 µg ml<sup>-1</sup> kanamycin and 12 µg ml<sup>-1</sup> tetracycline. *Listeria monocytogenes* Δ*Hly*Δ*fliC* lacking listeriolysin O (LLO) and flagellin expression were provided by D. Portnoy<sup>38</sup>.

**Treatment of macrophages and dendritic cells with viable and killed bacteria.** Macrophages were detached and re-plated 4 h before the experiment. BMDCs were re-plated immediately before addition of bacteria or soluble ligands. Unless stated otherwise, bacteria were used at a multiplicity of infection of 20. All experiments were carried out in antibiotic-free 'complete medium'. One hour after addition of bacteria, penicillin (100 µg ml<sup>-1</sup>) and streptomycin (100 µg ml<sup>-1</sup>) were added to the medium to kill any remaining extracellular bacteria. Alternatively, gentamicin sulphate (50 µg ml<sup>-1</sup>) was used. We also compared this approach to washing the cells and replacing the antibiotic-free medium with penicillin/streptomycin containing medium after 1 h and found no differences

with regards to the cellular responses measured. Supernatants were collected 24 h after the addition of the bacteria unless stated otherwise in the figure legends.

**Cytokine enzyme-linked immunosorbent assays.** Supernatants from cultured BMMs or BMDCs were collected at 24 h after stimulation or at the times indicated. Enzyme-linked immunosorbent assay (ELISA) antibody pairs used for IL-6, IL-1β and TNF-α were as listed below. All ELISA antibodies were used at 2 µg ml<sup>-1</sup> capture and 0.5 µg ml<sup>-1</sup> detection, with the exception of IL-6 capture, which was used at 1 µg ml<sup>-1</sup>. Detection antibodies were biotinylated and labelled by streptavidin-conjugated horseradish peroxidase (HRP), and visualized by the addition of o-phenylenediamine dihydrochloride (Sigma) (from tablets) or 3,3', 5,5'-tetramethylbenzidine solution (TMB, KPL). Colour development was stopped with 3 M H<sub>2</sub>SO<sub>4</sub> or TMB-Stop Solution (KPL), respectively. Recombinant cytokines served as standards and were purchased from Peprotech. Absorbances at 492 or 450 nm were measured, respectively, on a tunable microplate reader (VersaMax, Molecular Devices). Cytokine supernatant concentrations were calculated by extrapolating absorbance values from standard curves where known concentrations were plotted against absorbance using SoftMax Pro 5 software. Capture/detection antibody pairs were as follows. IL-6, MP5-20F3/MP5-32C11 (BD Pharmingen); IL-1β, B12/rabbit polyclonal antibody (eBioscience); TNF-α, TN3-19/rabbit polyclonal antibody (eBioscience). IFN-β production was measured from supernatants using the VeriKine Mouse IFN-Beta ELISA Kit (PBL Interferon source) following manufacturer's instructions.

**Anti-*E. coli* antibody ELISA.** 96-well microtitre plates were coated overnight with *E. coli* lysates (3 µg ml<sup>-1</sup>) that we generated from log-phase cultures of *thyA*<sup>-</sup> *E. coli*. Serum samples from immunized mice were serially diluted (12 dilutions) and incubated in the pre-coated plates for 12 h at 4 °C followed by washing and incubation with rabbit anti-mouse isotype-specific Ig-HRP (Southern Biotech) for 1 h. Bound rabbit anti-mouse Ig-HRP was visualized by the addition of o-phenylenediamine dihydrochloride (Sigma) from tablets, and the anti-*E. coli* antibody titres for each mouse were determined by absorbance readings at 490 nm.

**Measurement of inflammatory cell death.** Cell death of macrophages or BMDCs was measured using the Cytotox96 cytotoxicity assay (Promega) following manufacturer's instructions. The assay measures the release of lactate dehydrogenase (LDH) into the supernatant calculated as the percentage of total LDH content, measured from cellular lysates (100%). LDH released by unstimulated cells was used for background correction.

**Flow cytometric assessment of cell death.** Cells were stimulated overnight, stained for Annexin V/7AAD using the Annexin V-PE/7AAD Apoptosis Detection kit (BD Pharmingen), and analysed by flow cytometry (FACSCalibur, BD).

**Flow cytometric measurement of ROS production.** BMMs were loaded with the ROS indicator dye H2DCFDA (Molecular Probes/Invitrogen, 10 mM in PBS) for 30 min followed by a recovery time of 30 min in fresh pre-warmed 'complete medium'. BMMs were then stimulated with viable or heat killed *E. coli* for 60 min, washed and analysed by flow cytometry (FACSCalibur, BD).

**Western blots.** For detection of caspase-1, protein extracts were separated on 4–12% SDS-gradient gels (Invitrogen). For detection of all other proteins, samples were run on 10% SDS-polyacrylamide gels. Proteins were transferred to PVDF membranes (Millipore). Membranes were blocked with 5% milk in PBS and probed with the following antibodies: caspase-1 p10 (M-20)/rabbit polyclonal antibody, IkBα (C-21)/rabbit polyclonal antibody (both from Santa Cruz Biotechnologies), phospho-IRF3 (Ser 396)/rabbit polyclonal antibody, IRF3/rabbit polyclonal antibody, phospho-p38 MAPK (Thr 180/Tyr 182)/rabbit polyclonal antibody, p38 MAPK/rabbit polyclonal antibody (all from Cell Signalling Technology), α-tubulin (DM1A)/rabbit monoclonal antibody (Novus Biologicals).

**Real-time PCR.** Total RNA was isolated from macrophages using the RNeasy kit (Qiagen). Contaminating genomic DNA was removed by DNase digestion (DNase I, Promega). Reverse transcription was performed using Superscript III (Invitrogen) and cDNA was used for subsequent real-time PCR reactions. Quantitative real-time RT-PCR was conducted on an ABI Prism 7900 instrument using the Maxima SYBR green qPCR Master Mix (Fermentas) with the following primer pairs. β-Actin, FW 5'-GAAGTCCCTACCTCCCAA-3', RV 5'-GGC ATGGACGCGACCA-3'; *Il1b*, FW 5'-AAAGACGGCACACCCACCTGC-3', RV 5'-TGTCCTGACCACTGTTGTTTCCAG-3'; *Ifnb*, FW 5'-GCACTGGGT GGAAT-3', RV 5'-TTCTGAGGCATCAA-3'; *Nlrp3*, FW 5'-CGAGACCTCTG GGAAAAGCT-3', RV 5'-GCATACCATAGAGGAATGTGATGTACA-3'. All reactions were performed in duplicates and the samples were normalized to β-actin. 'Fold inductions' were calculated using the ΔΔC<sub>t</sub> method relative to unstimulated BMMs.

**Transcriptome analysis.** BMMs derived from wild-type or *Trif*<sup>-/-</sup> mice were stimulated with viable *E. coli* for 0, 1, 3 or 6 h and total RNA was extracted using the RNeasy kit (Qiagen). RNA from three independent experiments was used for

transcriptional analysis. RNA integrity was checked on an Agilent 2100 Bioanalyser (Agilent Technologies) with 6000 Nano Chips. RNA was judged as suitable only if samples showed intact bands of 18S and 28S ribosomal RNA subunits, displayed no chromosomal peaks or RNA degradation products, and had a RNA integrity number (RIN) above 8.0.

One-hundred nanograms of RNA were used for whole-transcript cDNA synthesis with the Ambion WT expression kit (Applied Biosystems). Hybridization, washing and scanning of an Affymetrix GeneChip Mouse Gene 1.1 ST 24-array plate was carried out according to standard Affymetrix protocols on a GeneTitan instrument (Affymetrix).

Packages from the Bioconductor project, integrated in an in-house developed management and analysis database for microarray experiments, were used for analysis of the scanned arrays<sup>39</sup>. Arrays were normalized using the Robust Multi-array Average method<sup>40,41</sup>. Probe sets were defined according to ref. 42. With this method probes are assigned to unique gene identifiers, in this case Entrez IDs. The probes on the Gene 1.1 ST arrays represent 19,807 genes that have at least 10 probes per identifier. For the analysis, only genes that had an intensity value of  $>20$  on at least two arrays were taken into account. In addition, the interquartile range of  $\log_2$  intensities had to be at least 0.25. These criteria were met by 9,921 genes. Changes in gene expression are represented as signal log ratios between treatment and control. Multiple Experiment Viewer software (MeV 4.6.1) was used to create heatmaps<sup>43,44</sup>. Genes were clustered by average linkage hierarchical clustering using Pearson correlation. Significantly regulated genes were identified by intensity-based moderated *t*-statistics<sup>45</sup>. Obtained *P*-values were corrected for multiple testing by a false discovery rate method<sup>46</sup>.

IFN-regulated genes were identified using the Interferome database (<http://www.interferome.org>)<sup>47</sup> and grouped in a heat map. Rel/NF- $\kappa$ B target genes were identified using another online database (<http://bioinfo.lifl.fr/NF-KB/>) which compiles Rel/NF- $\kappa$ B target genes identified by various groups<sup>48</sup> (<http://people.bu.edu/gilmore/nf-kb/index.html>). Inflammasome-related genes were compiled based on the current literature<sup>11,49</sup>.

**Measuring release from bacterial phagosomes.** Measurement of fluorescein-dextran (Fdx) release from macrophage phagosomes was performed using a modified method described previously<sup>16</sup>. BMMs were plated onto Mat-tek coverslip dishes (MatTek Corp.) and incubated overnight. BMMs were stimulated with viable or gentamicin-killed red fluorescent protein (RFP)-expressing *thyA*<sup>-</sup> *E. coli* in the presence of 0.167 mg ml<sup>-1</sup> Fdx in 200  $\mu$ l of medium. After 120 min of co-culture, additional Fdx and gentamicin containing medium was added to the coverslip dishes to prevent drying and to prevent bacterial overgrowth. Cells were imaged after 2, 4 and 8 h to measure release of Fdx. Microscopic imaging was performed on an IX70 inverted microscope (Olympus) equipped with an X-cite 120 metal halide light source (EXFO) and excitation and emission filter wheels. Phase contrast and two fluorescence images were acquired for each field of cells. The fluorescent images used the same emission settings, but used different excitation band-pass filters. Fdx fluorescence intensity using an excitation filter centred at 440 nm is relatively insensitive to pH, whereas fluorescence intensity using an excitation filter centred at 485 nm is very sensitive to pH. The ratio of fluorescence intensity at 485 nm divided by 440 nm was converted to into pH maps using calibration curves generated by imaging BMMs with Fdx-containing compartments at a series of fixed pH conditions. As described previously<sup>16</sup>, pixels with pH above 5.5 were designated as representing Fdx which has been released from endolysosomal compartments. The percentage of Fdx release was calculated by dividing the total intensity of pixels containing released Fdx by the total Fdx intensity for each cell.

**Infections and vaccinations.** For measurements of systemic cytokine levels, C57BL/6J wild-type, *Trif*<sup>-/-</sup>, *Asc*<sup>-/-</sup> or *Nlrp3*<sup>-/-</sup> mice were injected with  $1 \times 10^9$  viable or  $5 \times 10^9$  heat-killed *thyA*<sup>-</sup> *E. coli*, respectively. Blood samples were drawn 6 h after infection, and cytokine concentrations were measured by ELISA. For determination of bacterial clearance, we infected mice with  $1 \times 10^8$  viable replication-sufficient *E. coli* by intraperitoneal injection. Mice were monitored daily and moribund animals were killed according to humane criteria established and approved by our institutional IACUC committee. After 60 h, animals were killed and the spleens were explanted, homogenized, serially diluted and plated on LB-agar plates overnight followed by colony forming units (c.f.u.) counting.

For vaccinations, we followed a prime-boost regimen as shown in the schematic in Fig. 4h that was adopted from a previous study<sup>50</sup>. In brief, mice received an initial vaccination intraperitoneally with  $5 \times 10^7$  c.f.u. of viable or heat-killed *thyA*<sup>-</sup> *E. coli* or a combination of  $5 \times 10^7$  c.f.u. heat-killed *thyA*<sup>-</sup> *E. coli* and 30  $\mu$ g of purified *E. coli* total RNA, followed by two boosts ( $5 \times 10^6$  c.f.u.) after 10 and 20 days. Polyclonal class-specific anti-*E. coli* antibody production was measured in the serum after 25 days by ELISA.

**Bacterial RNA.** Total bacterial RNA was isolated from *thyA*<sup>-</sup> *E. coli* using the e.z.n.a. Bacterial RNA Kit (Omega Bio-Tek), following the manufacturer's instructions. Contaminating DNA was removed by DNase digestion (TURBO DNase,

Ambion/Applied Biosystems). Alternatively, total purified *E. coli* (DH5 $\alpha$ ) RNA was purchased from Ambion/Applied Biosystems, and similar results were obtained. Fractionation of bacterial RNA species was performed as follows. First, ribosomal 16S and 23S RNA (rRNA) was removed by a magnetic bead-based capture hybridization approach using the MICROBExpress kit (Ambion/Applied Biosystems). The enriched RNA was then separated into messenger RNA (mRNA) and small RNA (sRNA, including 5S rRNA) using the MEGAClear kit (Ambion/Applied Biosystems). All separated RNA fractions were precipitated with ammonium acetate and re-suspended in nuclease-free water. RNA concentration and purity were determined by measuring the absorbance at 260/280 and 260/230 nm. RNA preparations were further visualized by 1% agarose gel electrophoresis.

**In vitro RNA transcription.** The *E. coli* Gro operon encoding the bacterial chaperonins GroEL and GroES, the GTPase Era operon or the DNA polymerase III operon were PCR amplified from genomic DNA isolated from *thyA*<sup>-</sup> *E. coli* using primer pairs containing a T7 promoter sequence (T7) in either the FW or both FW and RV primer. Gro-FWT7 5'-TAATACGACTCACTATAGGGCACC AGCCGGGAAACCACG-3'; Gro-RVT7 5'-TAATACGACTCACTATAGGAA AAGAAAAACCCCCAGACAT-3'; Gro-RV 5'-AGATGACCAAAAGAAAAA CCCCAGACATT-3'; Era-FWT7 5'-TAATACGACTCACTATAGGGCATA TGAGCATCGATAAAAGTTAC-3'; Era-RV 5'-TTTAAAGATCGTCAACGT AACCGAG-3'; DNAPol-FWT7 5'-TAATACGACTCACTATAGGGATGTCTG AACACGTTTCGT-3'; DNAPol-RV 5'-AGTCAAATCCAGTTCACCTGC TCCGAA-3'.

PCR fragments were purified using the Nucleospin Extract II PCR purification kit (Macherey-Nagel), and used as DNA templates for *in vitro* transcription. *In vitro* transcription was performed using the MEGAScript kit T7 (Ambion/Applied Biosystems) following the manufacturer's instructions. DNA templates generated with Gro-FWT7 and Gro-RV primers only contained a T7 promoter site in the sense strand and yielded single-stranded RNA, whereas PCR templates generated with Gro-FWT7 and Gro-RVT7 primers contained T7 promoter sequences in both strands, allowing transcription of two complementary strands, yielding double-stranded RNA. For generation of 5'-capped RNA, m7G(5')ppp(5')G cap analogue (Ambion/Applied Biosystems) was included in the transcription reaction at a GTP:cap ratio of 1:4.

**RNA digestion, dephosphorylation and polyadenylation.** *In vitro*-transcribed Gro RNA, total *E. coli* RNA or *E. coli* mRNA were digested using RNase I (Promega) and RNase III (Ambion/Applied Biosystems). To remove 5'-triphosphates, RNA dephosphorylation was performed by incubating 10  $\mu$ g *in vitro*-transcribed RNA or total *E. coli* RNA or 1 mg of *E. coli* mRNA with 30 U of calf intestinal alkaline phosphatase (CIP, New England Biolabs) for 2 h at 37 °C, as described previously<sup>51</sup>. Polyadenylation of *in vitro*-transcribed and purified bacterial mRNA was performed using the poly(A) Tailing kit (Ambion) following the manufacturer's instructions.

**Transfection of macrophages and MEFs.** For direct cytosolic delivery of total purified *E. coli* RNA or *in vitro*-transcribed Gro RNA,  $5 \times 10^5$  BMMs or  $2 \times 10^5$  MEFs were transfected with 1 mg of RNA using 2  $\mu$ l of Lipofectamine 2000 (Invitrogen) in 24- or 12-well plates, respectively.

**Soluble ligands, inhibitors and other reagents.** Lipopolysaccharide (LPS) was purchased from Sigma (*E. coli* 055:B5, phenol extracted). Caspase inhibitors z-YVAD, z-IETD, Q-VD-OPH (all SM Biochemicals) were used at 50  $\mu$ M, and added 30 min before stimulation of cells.

**Statistical analysis.** Statistical significances were tested by an ANOVA Kruskal-Wallis test and Bonferroni-Dunn post hoc correction. Significances are represented in the figures as follows: \*, *P*  $\leq$  0.05; \*\*, *P*  $\leq$  0.01; \*\*\*, *P*  $\leq$  0.001. NS, not statistically significant; hash symbol, not detected.

- Torchinsky, M. B., Garaude, J., Martin, A. P. & Blander, J. M. Innate immune recognition of infected apoptotic cells directs TH17 cell differentiation. *Nature* **458**, 78–82 (2009).
- Blander, J. M. & Medzhitov, R. Regulation of phagosome maturation by signals from toll-like receptors. *Science* **304**, 1014–1018 (2004).
- Sutterwala, F. S. et al. Critical role for NALP3/CIA1/Cryopyrin in innate and adaptive immunity through its regulation of caspase-1. *Immunity* **24**, 317–327 (2006).
- Kuida, K. et al. Altered cytokine export and apoptosis in mice deficient in interleukin-1 $\beta$  converting enzyme. *Science* **267**, 2000–2003 (1995).
- Sato, M. et al. Distinct and essential roles of transcription factors IRF-3 and IRF-7 in response to viruses for IFN- $\alpha/\beta$  gene induction. *Immunity* **13**, 539–548 (2000).
- Maurelli, A. T., Baudry, B., d'Hauteville, H., Hale, T. L. & Sansonetti, P. J. Cloning of plasmid DNA sequences involved in invasion of HeLa cells by *Shigella flexneri*. *Infect. Immun.* **49**, 164–171 (1985).
- Haraga, A., Ohlson, M. B. & Miller, S. I. Salmonellae interplay with host cells. *Nature Rev. Microbiol.* **6**, 53–66 (2008).
- Schnupf, P. & Portnoy, D. A. Listeriolysin O: a phagosome-specific lysin. *Microbes Infect.* **9**, 1176–1187 (2007).
- Gentleman, R. C. et al. Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol.* **5**, R80 (2004).

40. Bolstad, B. M., Irizarry, R. A., Astrand, M. & Speed, T. P. A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics* **19**, 185–193 (2003).
41. Irizarry, R. A. *et al.* Summaries of Affymetrix GeneChip probe level data. *Nucleic Acids Res.* **31**, e15 (2003).
42. Dai, M. *et al.* Evolving gene/transcript definitions significantly alter the interpretation of GeneChip data. *Nucleic Acids Res.* **33**, e175 (2005).
43. Saeed, A. I. *et al.* TM4: a free, open-source system for microarray data management and analysis. *Biotechniques* **34**, 374–378 (2003).
44. Saeed, A. I. *et al.* TM4 microarray software suite. *Methods Enzymol.* **411**, 134–193 (2006).
45. Sartor, M. A. *et al.* Intensity-based hierarchical Bayes method improves testing for differentially expressed genes in microarray experiments. *BMC Bioinformatics* **7**, 538 (2006).
46. Storey, J. D. & Tibshirani, R. Statistical significance for genomewide studies. *Proc. Natl Acad. Sci. USA* **100**, 9440–9445 (2003).
47. Samarajiwa, S. A., Forster, S., Auchettl, K. & Hertzog, P. J. INTERFEROME: the database of interferon regulated genes. *Nucleic Acids Res.* **37**, D852–D857 (2009).
48. Pahl, H. L. Activators and target genes of Rel/NF- $\kappa$ B transcription factors. *Oncogene* **18**, 6853–6866 (1999).
49. Coll, R. C. & O'Neill, L. A. New insights into the regulation of signalling by toll-like receptors and nod-like receptors. *J. Innate Immun.* **2**, 406–421 (2010).
50. Lim, S. Y., Bauermeister, A., Kjønaas, R. A. & Ghosh, S. K. Phytol-based novel adjuvants in vaccine formulation: 2. Assessment of efficacy in the induction of protective immune responses to lethal bacterial infections in mice. *J. Immune Based Ther. Vaccines* **4**, 5 (2006).
51. Hornung, V. *et al.* 5'-Triphosphate RNA is the ligand for RIG-I. *Science* **314**, 994–997 (2006).



# Reprogramming transcription by distinct classes of enhancers functionally defined by eRNA

Dong Wang<sup>1\*</sup>, Ivan Garcia-Bassets<sup>2,3\*</sup>, Chris Benner<sup>1\*</sup>, Wenbo Li<sup>2</sup>, Xue Su<sup>2,4</sup>, Yiming Zhou<sup>5</sup>, Jinsong Qiu<sup>1</sup>, Wen Liu<sup>2,4</sup>, Minna U. Kaikkonen<sup>1</sup>, Kenneth A. Ohgi<sup>2</sup>, Christopher K. Glass<sup>1</sup>, Michael G. Rosenfeld<sup>2</sup> & Xiang-Dong Fu<sup>1</sup>

Mammalian genomes are populated with thousands of transcriptional enhancers that orchestrate cell-type-specific gene expression programs<sup>1–4</sup>, but how those enhancers are exploited to institute alternative, signal-dependent transcriptional responses remains poorly understood. Here we present evidence that cell-lineage-specific factors, such as FoxA1, can simultaneously facilitate and restrict key regulated transcription factors, exemplified by the androgen receptor (AR), to act on structurally and functionally distinct classes of enhancer. Consequently, FoxA1 downregulation, an unfavourable prognostic sign in certain advanced prostate tumours, triggers dramatic reprogramming of the hormonal response by causing a massive switch in AR binding to a distinct cohort of pre-established enhancers. These enhancers are functional, as evidenced by the production of enhancer-templated non-coding RNA (eRNA<sup>5</sup>) based on global nuclear run-on sequencing (GRO-seq) analysis<sup>6</sup>, with a unique class apparently requiring no nucleosome remodelling to induce specific enhancer–promoter looping and gene activation. GRO-seq data also suggest that liganded AR induces both transcription initiation and elongation. Together, these findings reveal a large repository of active enhancers that can be dynamically tuned to elicit alternative gene expression programs, which may underlie many sequential gene expression events in development, cell differentiation and disease progression.

The wide diversity of mammalian cells is determined by a large repertoire of constitutive and inducible genes, which are regulated by general and cell-type-specific transcription factors and cofactors through regulatory genomic elements<sup>7,8</sup>. Recent studies reveal that gene promoters are marked by tri-methylated H3K4 (H3K4me3) and distal regulatory elements are often associated with mono-methylated H3K4 (H3K4me1)<sup>1,2</sup>. Because these H3K4me1-positive, H3K4me3-negative regions exhibit striking cell-type specificity<sup>1,2</sup>, we used this signature to characterize potential enhancers in prostatic LNCaP cells in which one of key regulatory transcriptional programs is mediated by the AR. We identified by chromatin immunoprecipitation (ChIP)-sequencing 14,283 H3K4me3-marked and 51,544 H3K4me1-marked loci in androgen (5 $\alpha$ -dihydrotestosterone, (DHT))-treated LNCaP cells, among which 43,565 loci are uniquely marked by H3K4me1, largely localized distal to annotated transcriptional start sites (TSSs) (94%), and associated with other marks linked to enhancer activities (Fig. 1a).

*De novo* DNA motif analysis revealed several highly enriched motifs, particularly the forkhead motif (Fig. 1b). Using a specific antibody against FoxA1, a major FOX family member expressed in LNCaP cells and normal prostate gland<sup>9–11</sup> (Supplementary Fig. 1), we identified 33,426 FoxA1-bound sites, which extensively overlap with distal H3K4me1-marked regions (Fig. 1c and Supplementary Fig. 2a; see on *KLK3* enhancer<sup>12</sup> in Supplementary Fig. 2b). RNA profiling supports the functional relevance of these FoxA1/H3K4me1 loci, as genes responsive to *FOXA1* short interfering RNA (siRNA) are

located more proximally to FoxA1/H3K4me1-marked loci than non-responsive genes (Fig. 1d and Supplementary Fig. 3).

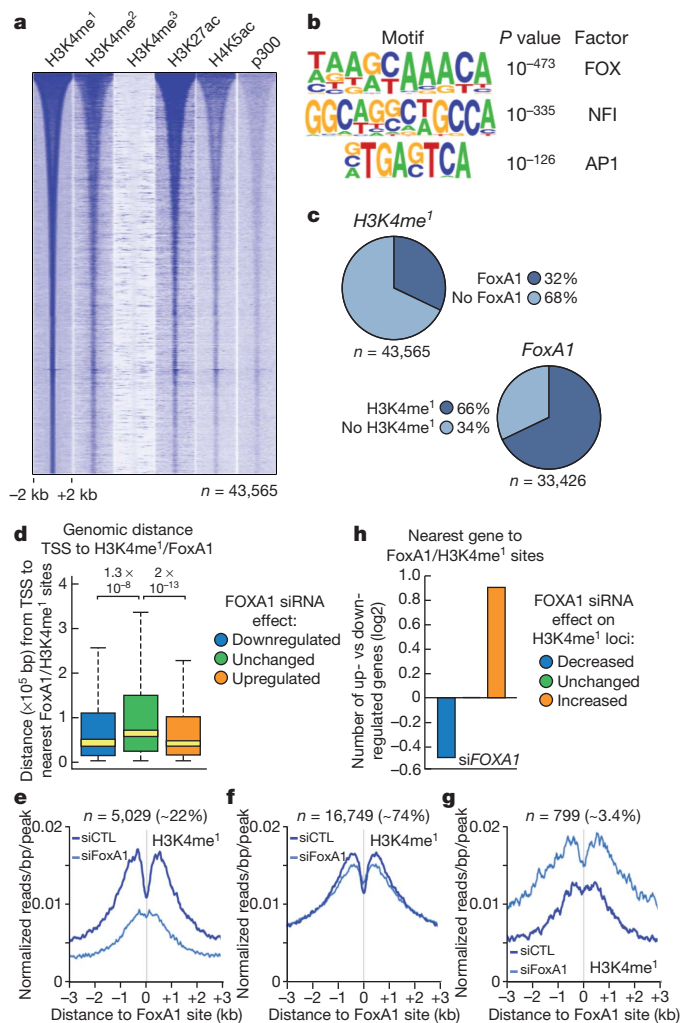
FoxA1 has been characterized as a ‘pioneer’ factor to facilitate DNA binding by other sequence-specific transcription factors<sup>9,13–16</sup> and ‘translate’ H3K4me1/me2 into AR-mediated gene expression<sup>9</sup>. Comparing the profile of H3K4me1 and H3K27ac before and after *FOXA1* knock-down, we detected three classes of FoxA1-binding sites based on the H3K4me1 signal exhibiting reduced (~22%), relatively unaffected (~74%) or even increased (~3.4%) levels over candidate enhancers (Fig. 1e–g and Supplementary Fig. 4). RNA profiling analysis agrees with the functional significance of these selective FoxA1 effects, revealing more downregulated genes in the first class, roughly equal numbers of up- or downregulated genes in the second and more upregulated genes in the third (Fig. 1h), suggesting a contribution of FoxA1 to ‘writing’ and ‘reading’ the ‘histone code’ on different enhancer cohorts, in line with its critical function in prostate gland development<sup>10,11</sup>.

The rationale for our experimental strategy to use RNA interference (RNAi) to study FoxA1-regulated enhancer network is the association of decreased *FOXA1* expression with castration-resistant, poor prognostic prostate tumours (Supplementary Fig. 5). In LNCaP cells, *FOXA1* RNAi enhanced cell entrance to S phase with reduced hormone (Fig. 2a). To understand the mechanistic basis for elevated hormone responsiveness, we mapped AR-binding sites, identifying 3,115 high confident loci with approximately 65% co-incident with H3K4me1. *De novo* motif analysis revealed highly enriched elements for both AR and FoxA1, including a composite motif consisting of a FOX motif and AR regulatory element (ARE) half site, suggesting ternary complex formation on these sites (Fig. 2b). Indeed, 1,684 AR-bound loci (54% of total) are co-occupied by FoxA1 in DHT-treated LNCaP cells and FoxA1 appears to bind to most of these sites (~70%) before hormone treatment (Supplementary Fig. 6).

The conundrum is that, although FoxA1 is known to facilitate AR binding on several DHT-responsive genes<sup>9</sup>, *FOXA1* RNAi actually markedly elevated, rather than diminished, the DHT response (Fig. 2a). We found that approximately 60% of the original AR binding events were ‘expectedly’ lost in response to *FOXA1* RNAi, which we refer to as the ‘lost’ AR program (Fig. 2c, d). We refer to the remaining approximate 40% of AR binding events as the ‘conserved’ AR program, which often exhibited enhanced AR binding. Strikingly, we detected a massive gain of 10,869 new AR binding loci, referred to as the ‘gained’ AR program (Fig. 2c, d). We extensively validated each of these AR programs by conventional ChIP–quantitative PCR (qPCR) (Fig. 2e). This induced AR reprogramming appears to be qualitatively and quantitatively distinct from reported AR re-targeting on androgen-resistant LNCaP-abl cells compared with parental LNCaP cells<sup>17</sup> and is in sharp contrast to FoxA1-dependent genomic targeting of the oestrogen receptor- $\alpha$  (ER- $\alpha$ ) in breast cancer MCF7 cells<sup>18</sup>. In concert with such massive AR reprogramming, we observed corresponding

<sup>1</sup>Department of Cellular and Molecular Medicine, School of Medicine, University of California, San Diego, 9500 Gilman Drive, La Jolla, California 92093-0651, USA. <sup>2</sup>Department of Medicine, Howard Hughes Medical Institute, School of Medicine, University of California, San Diego, 9500 Gilman Drive, La Jolla, California 92093-0651, USA. <sup>3</sup>Department of Medicine, Division of Endocrinology and Metabolism, School of Medicine, University of California, San Diego, 9500 Gilman Drive, La Jolla, California 92093-0651, USA. <sup>4</sup>Graduate Program in Biology, School of Medicine, University of California, San Diego, 9500 Gilman Drive, La Jolla, California 92093-0651, USA. <sup>5</sup>Digomics LLC, PO Box 30, Malden, Massachusetts 02148, USA.

\*These authors contributed equally to this work.



**Figure 1 | FoxA1 contributes to the enhancer code in prostate cancer cells.**

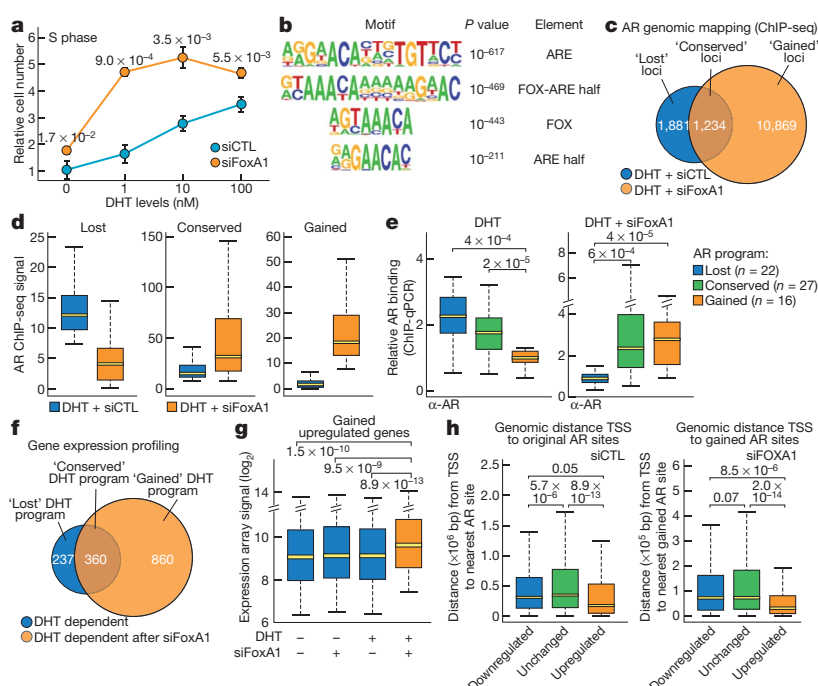
**a**, Distribution of histone marks within  $\pm 2$ -kb windows around distinct genomic regions ( $n = 43,565$ ) marked by H3K4me<sup>1</sup>, but not H3K4me<sup>3</sup>, in androgen (DHT)-stimulated LNCaP cells. The ChIP-seq data sets for H3K4me<sup>1</sup>, H3K4me<sup>2</sup>, H3K4me<sup>3</sup>, H3K27ac, H4K5ac and p300 were each aligned with respect to the centre of the H3K4me<sup>1</sup> signal and sorted by the length of H3K4me<sup>1</sup>-marked regions. **b**, Top-enriched DNA motifs with significant  $P$  values and prospective families of DNA binding transcription factors identified by *de novo* motif analysis of non-promoter regions marked by H3K4me<sup>1</sup>. **c**, Percentage of H3K4me<sup>1</sup>-marked regions that show FoxA1 binding events (top panel) and percentage of FoxA1-binding sites that are marked by H3K4me<sup>1</sup> (bottom panel). Note that H3K4me<sup>1</sup>-marked regions tend to be broad, but FoxA1-binding sites are discrete; as a result, many H3K4me<sup>1</sup>-positive regions may contain more than one FoxA1-binding site. **d**, Genomic distance from FoxA1/H3K4me<sup>1</sup>-positive loci to the nearest TSS of genes in response to FOXA1 knockdown. Outliers were omitted from box plots.  $P$  values indicate the significance in pair-wise comparisons. **e–g**, Three classes of FoxA1/H3K4me<sup>1</sup>-positive loci according to the response in levels of H3K4me<sup>1</sup> to FOXA1 knockdown: greater than 1.5-fold decrease (**e**), no significant change (**f**) and greater than 1.5-fold increase (**g**). **h**, Ratio (log<sub>2</sub>) of up- and downregulated genes in each H3K4me<sup>1</sup>-responsive category in **e–g**. CTL, control.

changes in gene expression in each of three AR programs (Fig. 2f, g and Supplementary Fig. 7). The newly induced AR expression program is also linked to AR binding events (Fig. 2h), suggesting a direct gain-of-function on DHT-responsive genes, as illustrated on SOX9 and other genes (Supplementary Fig. 8), which have been previously documented to play critical roles in cancer progression<sup>19,20</sup>. Because we also observed an approximate threefold elevation of AR expression in FOXA1 RNAi-treated cells (Supplementary Fig. 9a), we tested the possibility that increased AR expression might trigger these effects. We found that AR overexpression alone was insufficient to induce AR reprogramming (Supplementary Fig. 9b).

To explore the mechanism for AR reprogramming, we determined FoxA1 binding on different AR programs. We found that the gained AR program is largely devoid of FoxA1, whereas FoxA1 is present in more than half of the lost and conserved AR programs (Supplementary Fig. 10). This raises the possibility that FoxA1 may facilitate AR binding to its original binding program, but trans-repress AR from binding to other genomic regions that lack FoxA1-binding sites in the gained program, a strategy frequently used by other transcription activators<sup>21</sup>.

**Figure 2 | AR reprogramming and induced alternative hormonal response.**

**a**, FOXA1 siRNA-induced cell progression to S phase. Relative numbers of propidium-iodide-labelled cells in S-phase at different DHT concentrations were determined by FACS. The  $P$  value for the difference detected at each hormonal level is indicated; mean  $\pm$  s.e.m. is based on four independent experiments. **b**, Top-enriched motifs associated with AR-occupied loci ( $n = 3,115$ ). **c**, Comparison between genome-wide AR binding programs before and after FOXA1 knockdown in DHT-treated LNCaP cells. **d**, Quantitative levels of AR binding in the 'lost', 'conserved' and 'gained' programs. Outliers were omitted from box plots. **e**, ChIP-qPCR-validated AR binding events on randomly selected loci from the lost ( $n = 22$ ), conserved ( $n = 27$ ) and gained ( $n = 16$ ) programs. **f**, Microarray analysis of DHT-induced genes before and after FOXA1 knockdown. **g**, Quantitative analysis of gained androgen upregulated genes based on microarray analysis in **f**. Outliers were omitted from box plots. **h**, Genomic distance of androgen-responsive genes from TSS to the nearest AR-binding site in the original and gained AR-binding programs.

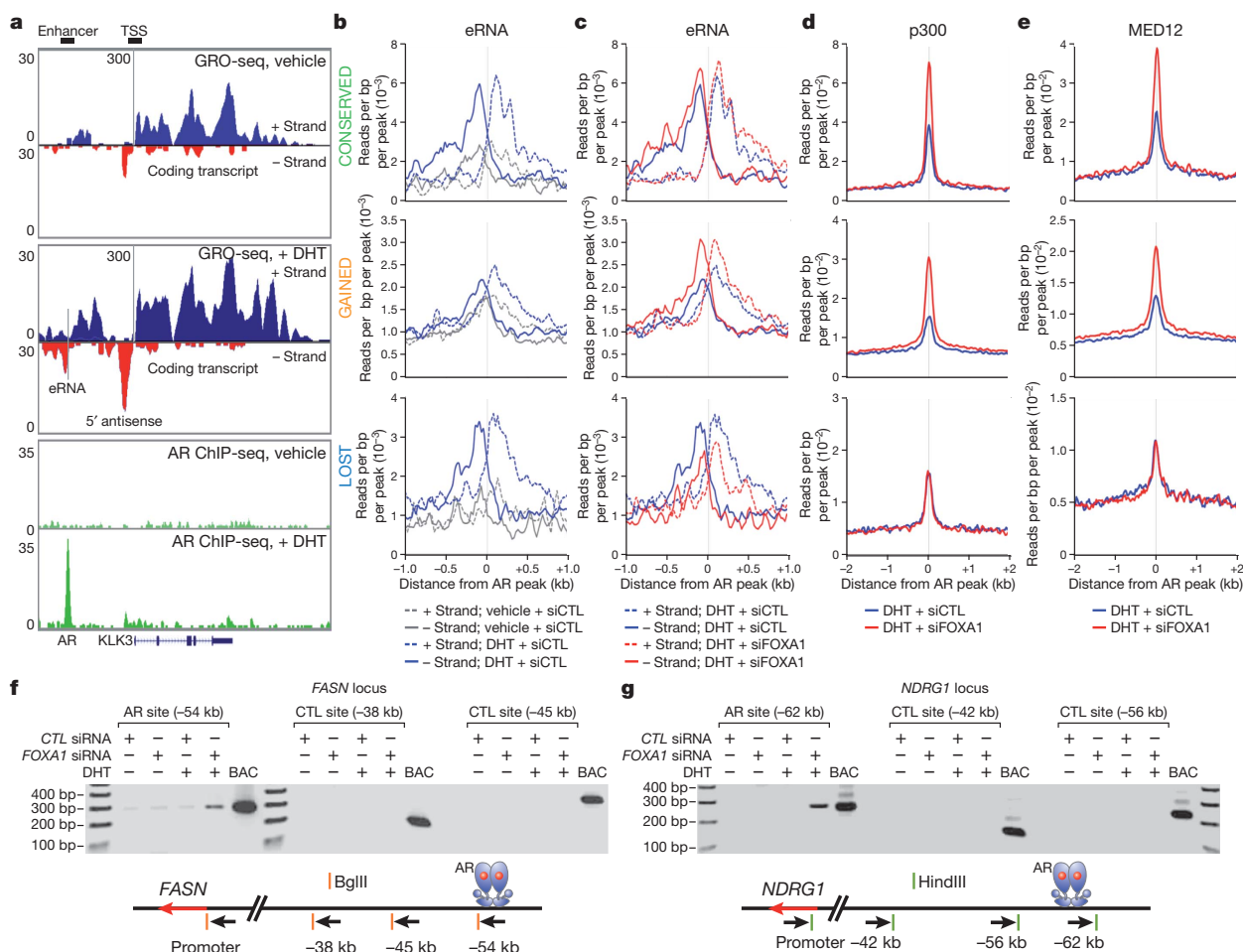


Indeed, as previously reported<sup>22</sup>, FoxA1 overexpression squelched ARE-driven transcription in transfected HEK293 cells (Supplementary Fig. 11), which is consistent with the ability of AR to interact with FoxA1 directly<sup>23</sup>. This mechanism appears to be exploited during tumour progression because an AR mutation identified in advanced prostate tumours lacks part of the hinge domain important for interactions with FoxA1, its ability to interact with FoxA1, and became resistant to FoxA1-mediated trans-repression (Supplementary Fig. 11b, c). Furthermore, our functional analysis indicates that the missing AR ligand-binding domain also contributes to AR:FoxA1 interactions (Supplementary Fig. 12). Interestingly, similar AR truncations have also been reported to result from alternative splicing, gene rearrangement and/or calpain-mediated cleavage (Supplementary Fig. 13). Based on these findings, we propose that FoxA1 regulates AR genomic targeting by simultaneously anchoring AR to cognate loci and restricting AR from other ARE-containing loci in the human genome.

To understand how reprogrammed AR binding is translated to altered hormonal response, we took advantage of the recently established GRO-seq<sup>6</sup> to detect the functional relationship between AR binding and hormone-induced gene expression. This powerful genome-wide interrogation of ongoing transcription detected a broad scope of nascent RNAs. We uncovered 28,318 transcripts with 15,656 annotated and 12,662 unannotated transcripts, among which 450 coding and 347 unannotated transcripts were induced more than 1.5-fold

with even just 1 h DHT treatment (Supplementary Fig. 14). The TSSs of GRO-seq defined transcripts are typically marked by H3K4me3 and H3K27ac (Supplementary Fig. 15a, b). Importantly, GRO-seq also detected non-coding RNAs from a subset of H3K4me1-positive, H3K4me3-negative regions (Supplementary Fig. 15c). As illustrated on the enhancer of the *KLK3* transcription unit (Fig. 3a), these eRNAs are largely symmetrical and bidirectional (see additional examples on other well-known hormone regulated genes, such as *PMEPA1* and *KLK2* in Supplementary Fig. 16). Interestingly, we often detected a large amount of nascent RNA before DHT treatment, particularly near their TSSs (for example, *KLK3*); DHT not only enhanced the expression of these nascent RNAs, but also allowed the extension of transcription towards the end of the gene (Fig. 3a and Supplementary Fig. 16). We estimated that approximately 79% of the transcription units induced by liganded AR are regulated at the level of transcriptional initiation, whereas approximately 21% appear to be primarily regulated at the level of elongation (Supplementary Fig. 17).

The ability to detect regulated eRNA expression allowed us to analyse different AR programs during transcriptional reprogramming. In the presence of FoxA1, DHT enhanced eRNA expression from AR-bound enhancers in both the lost and conserved AR programs. In contrast, a basal level of eRNAs was detectable on the gained program, but was independent of the hormone treatment, indicating that these are pre-established enhancers (Fig. 3b). In response to *FOXA1* RNAi,



**Figure 3 | Transcriptional response on individual enhancer programs to *FOXA1* downregulation.** **a**, Display of nascent RNA detected by GRO-seq on the *KLK3* locus. The DHT-induced AR binding is shown at bottom as a reference. **b, c**, Induction of eRNA by DHT (**b**) or *FOXA1* knockdown in DHT-treated LNCaP cells (**c**). The eRNA levels under different conditions (indicated at bottom) are separately displayed on three AR-binding programs. **d, e**, Effects

of FoxA1 on binding of p300 (**d**) and Med12 (**e**) in each AR program in DHT-treated LNCaP cells. **f, g**, Long-distance interaction between gene promoter and AR-bound site was determined by the 3C assay on two representative gene loci selected from the conserved and gained AR programs. Negative controls at shorter distances and a positive control with the corresponding BAC in the region are included in each case.



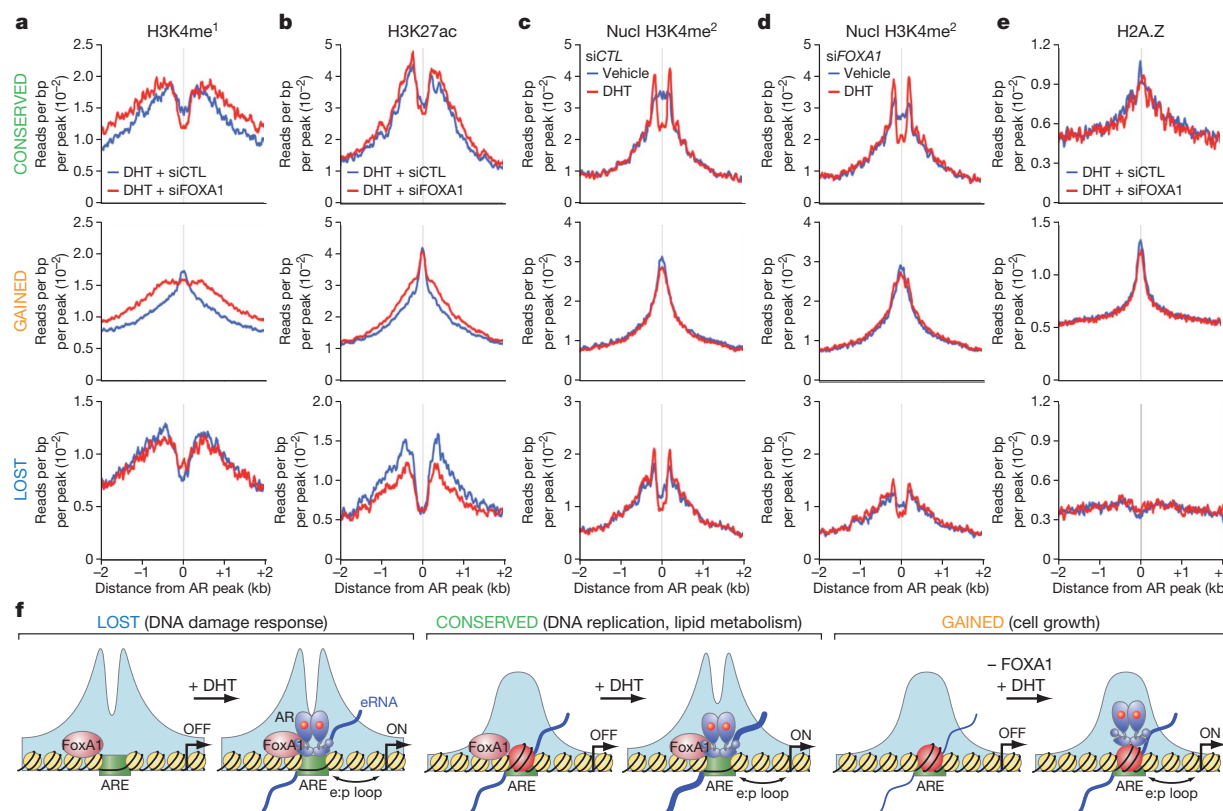
the expression of eRNAs was diminished from the lost program, but modestly or dramatically enhanced from the conserved and gained programs, respectively (Fig. 3c). The DHT-induced nascent transcripts (detected by GRO-seq) and steady-state RNAs (detected by microarrays) best predict direct target genes by liganded AR, as they show the shortest distance (<50 kilobases (kb)) to nearby AR-binding sites compared with genes identified by either criterion alone (Supplementary Fig. 18), indicating that AR-activated enhancers marked by increased eRNA are responsible for activation of nearby coding transcription units.

In concert with differential eRNA expression, we also observed corresponding changes in levels of another mark in the final step of enhancer activation<sup>4</sup>, specifically p300, on both conserved and gained AR programs (Fig. 3d). Interestingly, enhancers in the lost AR program continued to exhibit significant p300 binding, even after AR binding and eRNA expression were diminished in *FOXA1* knockdown cells (Fig. 3c, d). The transcription mediator Med12 has recently been suggested to mediate enhancer–promoter looping<sup>24</sup>. We tested Med12 binding on individual AR programs, finding that it exhibited an identical binding pattern to p300 (Fig. 3e). Enhanced Med12 binding on the conserved and gained programs after *FOXA1* knockdown suggests elevated or newly activated enhancer–promoter interactions. This was demonstrated by the 3C assay on two representative genes where *FOXA1* knockdown either enhanced (on the *FASN* locus from the conserved AR program) or created new (on the *NDRG1* locus in the gained AR program) long-range interactions between AR-bound enhancers and specific gene promoters in DHT-treated cells (Fig. 3f,

g and Supplementary Fig. 19). These data strongly suggest that the induction of eRNAs, rather than binding of either p300 or Med12, is the most precise mark of the final, functional looping between an activated enhancer and its regulated gene promoter.

Addressing the structural basis for different functional classes of AR enhancers, we note that the distinct profiles of H3K4me1 and H3K27ac on the lost, conserved and gained AR programs and *FOXA1* RNAi had little effect on these profiles (Fig. 4a, b and Supplementary Fig. 20). The histone marks H3K4me1 and H3K27ac around the lost and conserved AR programs exhibit a bimodal distribution, which is particularly pronounced on the lost program (Fig. 4a, bottom panel). The DNA-binding sites in the lost AR program are actually significantly less enriched in canonical AREs, which may render AR binding on these sites particularly dependent on FoxA1, whereas both the conserved and gained AR programs are associated with nearly perfect palindromic, canonical AREs (Supplementary Fig. 21), explaining why AR is able to target those sites in a FoxA1-independent manner. Strikingly, the gained AR-binding sites are coincident with sharp H3K4me1 and H3K27ac peaks (Fig. 4a, b, middle panels), suggesting a distinct nucleosome architecture underlying the gained AR program.

A recent study has suggested that AR binding leads to dynamic dismissal of a central, H2A.Z-containing nucleosome, being replaced by two flanking H3K4me2-marked nucleosomes<sup>25</sup>. We found that the lost AR program was largely devoid of a ‘central’ nucleosome even before AR binding (Fig. 4c, bottom panel). The conserved AR program exhibited DHT-induced switch from the central H3K4me2-marked nucleosome to two flanking H3K4me2-marked nucleosomes, which



**Figure 4 | Distinct classes of AR enhancers in the human genome.**

**a, b,** Profiles of H3K4me1 (**a**) and H3K27ac (**b**) associated with the lost (bottom panels), conserved (top panels) and gained (middle panels) AR programs in DHT-treated LNCaP cells in response to *FOXA1* knockdown. **c, d,** Profiles of H3K4me2 around AR binding loci at the nucleosomal resolution in response to DHT stimulation in control siRNA-treated (**c**) or *FOXA1* siRNA-treated (**d**) LNCaP cells. **e,** Profiles of the histone variant H2A.Z on the three different AR programs. **f,** Model for FoxA1-mediated AR targeting and reprogramming in LNCaP cells. In class I (the lost AR program), FoxA1 licenses liganded AR to

bind to ARE in relatively nucleosome-free regions. AR binding does not induce nucleosome remodelling in this class of enhancers. In class II (the conserved AR program), AR binds independently of FoxA1 to ARE, inducing nucleosome remodelling. In class III (the gained AR program), FoxA1 restricts AR binding, despite the presence of strong AREs. Although pre-established, these gained loci exhibit a strong central nucleosomes and are associated with H2A.Z, which is not affected by AR binding. *FOXA1* knockdown converted these sites to androgen-responsive sites. In all these three classes, eRNAs were generated or increased after AR binding. e:p, enhancer:promoter.

is largely independent of FoxA1 (Fig. 4c, d, top panels). The gained program showed a strong H3K4me2-marked central nucleosome both before and after AR binding (Fig. 4c, d, middle panel). Thus, this gained AR program represents a new type of enhancer topography that requires no nucleosome remodelling for enhancer recognition and subsequent enhancer–promoter interactions. H2A.Z is prevalently associated with the gained AR program, modestly with the conserved AR program and absent in the lost AR program (Fig. 4e). Together, these findings establish distinct chromatin structures underlying functionally distinct classes of AR enhancer.

In summary, our findings imply a general principle for establishing cell-type-specific transcription programs. Cell-lineage-specific factors (such as FoxA1) coupled with other general transcriptional factors ‘create’ a cell-type-specific enhancer network, allowing other regulated factors (such as AR) to ‘activate’ these pre-established enhancers (Fig. 4f). The enhancer activation process is tightly linked to eRNA production, which appear to serve as a more robust indicator of enhancer activities than any enhancer-bound transcription activators or chromatin marks. On the current biology model, AR reprogramming dramatically altered the androgen-responsive pathway, which, according to GO analysis (Supplementary Fig. 22 and Fig. 23), may contribute to enhanced cell growth and the establishment of an appropriate microenvironment in advanced prostate cancer<sup>26–28</sup>. Together, these findings provide a conceptual framework to understand complex gene-expression switching events, as occurs during disease progression and development.

## METHODS SUMMARY

Experiments were performed on LNCaP cells, LNCaP-AR cells (gift of C. Sawyers) and HEK293 cells. ChIPs were done as previously described<sup>29</sup> and GRO was performed as described<sup>6,30</sup>. Control siRNA was purchased from Qiagen (1027280). FOXA1 siRNA 1 (M-101319) and 2 (sense 5'-GAGAGAAAAAUCACAGC-3'; antisense 5'-GCUGUUGAUUUUUUCUCUC-3')<sup>9</sup> were purchased from or synthesized by Dharmacon.

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Received 30 July 2010; accepted 15 March 2011.**

**Published online 15 May 2011.**

- Heintzman, N. D. *et al.* Histone modifications at human enhancers reflect global cell-type-specific gene expression. *Nature* **459**, 108–112 (2009).
- Heintzman, N. D. *et al.* Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nature Genet.* **39**, 311–318 (2007).
- Pennacchio, L. A. *et al.* *In vivo* enhancer analysis of human conserved non-coding sequences. *Nature* **444**, 499–502 (2006).
- Visel, A. *et al.* ChIP-seq accurately predicts tissue-specific activity of enhancers. *Nature* **457**, 854–858 (2009).
- Kim, T. K. *et al.* Widespread transcription at neuronal activity-regulated enhancers. *Nature* **465**, 182–187 (2010).
- Core, L. J., Waterfall, J. J. & Lis, J. T. Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. *Science* **322**, 1845–1848 (2008).
- Lee, T. I. & Young, R. A. Transcription of eukaryotic protein-coding genes. *Annu. Rev. Genet.* **34**, 77–137 (2000).
- Rosenfeld, M. G., Lunyak, V. V. & Glass, C. K. Sensors and signals: a coactivator/corepressor/epigenetic code for integrating signal-dependent programs of transcriptional response. *Genes Dev.* **20**, 1405–1428 (2006).
- Lupien, M. *et al.* FoxA1 translates epigenetic signatures into enhancer-driven lineage-specific transcription. *Cell* **132**, 958–970 (2008).
- Gao, N. *et al.* Forkhead box A1 regulates prostate ductal morphogenesis and promotes epithelial cell maturation. *Development* **132**, 3431–3443 (2005).
- Mirosevich, J., Gao, N. & Matusik, R. J. Expression of Foxa transcription factors in the developing and adult murine prostate. *Prostate* **62**, 339–352 (2005).

- Kang, Z., Janne, O. A. & Palvimo, J. J. Coregulator recruitment and histone modifications in transcriptional regulation by the androgen receptor. *Mol. Endocrinol.* **18**, 2633–2648 (2004).
- Shim, E. Y., Woodcock, C. & Zaret, K. S. Nucleosome positioning by the winged helix transcription factor HNF3. *Genes Dev.* **12**, 5–10 (1998).
- Holmqvist, P. H., Belikov, S., Zaret, K. S. & Wrangé, O. FoxA1 binding to the MMTV LTR modulates chromatin structure and transcription. *Exp. Cell Res.* **304**, 593–603 (2005).
- Crowe, A. J. *et al.* Hepatocyte nuclear factor 3 relieves chromatin-mediated repression of the  $\alpha$ -fetoprotein gene. *J. Biol. Chem.* **274**, 25113–25120 (1999).
- Cirillo, L. A. *et al.* Opening of compacted chromatin by early developmental transcription factors HNF3 (FoxA) and GATA-4. *Mol. Cell* **9**, 279–289 (2002).
- Wang, Q. *et al.* Androgen receptor regulates a distinct transcription program in androgen-independent prostate cancer. *Cell* **138**, 245–256 (2009).
- Hurtado, A., Holmes, K. A., Ross-Innes, C. S., Schmidt, D. & Carroll, J. S. FOXA1 is a key determinant of estrogen receptor function and endocrine response. *Nature Genet.* **43**, 27–33 (2011).
- Thomsen, M. K. *et al.* SOX9 elevation in the prostate promotes proliferation and cooperates with PTEN loss to drive tumor formation. *Cancer Res.* **70**, 979–987 (2010).
- Wang, H. *et al.* SOX9 is expressed in human fetal prostate epithelium and enhances prostate cancer invasion. *Cancer Res.* **68**, 1625–1630 (2008).
- Ogawa, S. *et al.* Molecular determinants of crosstalk between nuclear receptors and toll-like receptors. *Cell* **122**, 707–721 (2005).
- Lee, H. J., Hwang, M., Chattopadhyay, S., Choi, H. S. & Lee, K. Hepatocyte nuclear factor-3 $\alpha$  (HNF-3 $\alpha$ ) negatively regulates androgen receptor transactivation in prostate cancer cells. *Biochem. Biophys. Res. Commun.* **367**, 481–486 (2008).
- Gao, N. *et al.* The role of hepatocyte nuclear factor-3 $\alpha$  (Forkhead Box A1) and androgen receptor in transcriptional regulation of prostatic genes. *Mol. Endocrinol.* **17**, 1484–1507 (2003).
- Kagey, M. H. *et al.* Mediator and cohesin connect gene expression and chromatin architecture. *Nature* **467**, 430–435 (2010).
- He, H. H. *et al.* Nucleosome dynamics define transcriptional enhancers. *Nature Genet.* **42**, 343–347 (2010).
- Laitinen, S., Karhu, R., Sawyers, C. L., Vessella, R. L. & Visakorpi, T. Chromosomal aberrations in prostate cancer xenografts detected by comparative genomic hybridization. *Genes Chromosom. Cancer* **35**, 66–73 (2002).
- Legrier, M. E. *et al.* Hormone escape is associated with genomic instability in a human prostate cancer model. *Int. J. Cancer* **124**, 1103–1111 (2009).
- Wolf, M. *et al.* High-resolution analysis of gene copy number alterations in human prostate cancer using CGH on cDNA microarrays: impact of copy number on gene expression. *Neoplasia* **6**, 240–247 (2004).
- Garcia-Bassets, I. *et al.* Histone methylation-dependent mechanisms impose ligand dependency for gene activation by nuclear receptors. *Cell* **128**, 505–518 (2007).
- Ingolia, N. T., Ghaemmaghami, S., Newman, J. R. & Weissman, J. S. Genome-wide analysis *in vivo* of translation with nucleotide resolution using ribosome profiling. *Science* **324**, 218–223 (2009).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We are grateful to H.-J. Kung for providing the PC3-AR cell lines, C. Sawyers for providing the LNCaP-AR cell line, and J. Hightower, D. Benson and M. Fisher for assistance with figure and manuscript preparation. This work was supported by grants from US Army Medical Research and Materiel Command (grant W81XWH-05-1-0100) to I.G.-B., the National Institutes of Health (DK01847, DK37949, DK074868, NS34934), Department of Defense and the National Cancer Institute (CA97134) to M.G.R.; the National Institutes of Health (GM049369 and HG004659) to X.-D.F.; and the Prostate Cancer Foundation to M.G.R. and X.-D.F. M.G.R. is an investigator of the Howard Hughes Medical Institute.

**Author Contributions** D.W., I.G.-B., M.G.R. and X.-D.F. designed the experiments; D.W. and I.G.B. performed most of them. C.B. performed all computational analyses. W.L. and M.K. performed GRO-seq; X.S. performed the 3C assay; Y.Z. analysed human samples; and J.Q. performed proliferation assays. K.A.O. and W.L. generated AR constructs. I.G.-B., D.W., M.G.R. and X.-D.F. wrote the manuscript with contributions from C.B. and C.K.G.

**Author Information** High-throughput data are deposited in Gene Expression Omnibus under accession number GSE27823 for all ChIP-seq and GRO-seq experiments and GSE27682 for microarray data. Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at [www.nature.com/nature](http://www.nature.com/nature). Correspondence and requests for materials should be addressed to X.-D.F. ([xdfu@ucsd.edu](mailto:xdfu@ucsd.edu)), M.G.R. ([mrosenfeld@ucsd.edu](mailto:mrosenfeld@ucsd.edu)) or I.G.-B. ([ibassets@ucsd.edu](mailto:ibassets@ucsd.edu)).

## METHODS

**Antibodies.** Specific antibodies were purchased from the following commercial sources: anti-FoxA1 (ab5089), anti-H3K4me1 (ab8899), anti-H3K27ac (ab4729), anti-H3K36me3 (ab9050) and H2A.Z (ab4174) from Abcam; anti-AR (N-20), anti-FoxA1 (C-20) and p300 (C-20, gift of B. Ren) from Santa Cruz Biotechnology; anti-H3K4me2 (07-030), anti-H3K4me3 (07-473), anti-H4K5ac (07-327) and anti-H3K27me3 (07-449) from Upstate Biotechnology; anti-Med12 (A300-774A) from Bethyl Laboratories; and anti-beta Actin (AC74) from Sigma.

**siRNA transfection.** One day before transfection, LNCaP cells were seeded in RPMI 1640 medium with 10% FBS. Six hours after siRNA transfection (20 pmol ml<sup>-1</sup>) with Lipofectamine 2000 (Invitrogen), cells were washed twice with PBS and then maintained in hormone-depleted phenol-free RPMI 1640 media. For gene expression profiling and western blotting, cells were cultured for 3 days after transfection and then treated with DHT for 20 h; for ChIP-qPCR and ChIP-seq, cells were cultured for 4 days after transfection and then treated with DHT for 1 h. ChIP-seq analysis at the nucleosome resolution was based on cells treated with DHT for 4 h.

**3C assay.** Cells were crosslinked with 1% formaldehyde for 20 min at room temperature and processed according to the standard 3C protocol<sup>31</sup>. For the study on the *FASN* locus, fixed chromatin from  $5 \times 10^6$  cells was digested with 400 units of BglII and EcoRI (NEB). For the *NDRG1* locus, fixed chromatin from  $5 \times 10^6$  cells was digested with 400 units of HindIII (NEB). Ligation was done with 800 units of T4 DNA ligase (NEB) for 4 h. The 3C product was quantified by qPCR after diluting the template tenfold compared with purified genomic DNA of known concentration. For each semi-quantitative PCR, the amount of template was titrated to determine the linear range in which the PCR product was amplified. PCR primers were designed next to BglII and HindIII restriction sites, respectively, for the *FASN* (all in minus strand) promoter (5'-AAGCTGTGAGTCAGCATGGTAG-3') and three upstream sites (-38kb, 5'-TGTCTTCTGATGTGTCTGCTTAGAG-3'; -45kb, 5'-AATCTGCTCAGGAATCTGTATGT-3'; -54kb, 5'-GGACACTACTGCTTTTCTGTG-3') and for the *NDRG1* (all in plus strand) promoter (5'-ATAGGTTCTGCCTTATTAGGG-3') and three upstream sites (-42kb, 5'-ATAGAGTTAGAGAAACGGAGGCAGT-3'; -56kb, 5'-GCCGTGAAGATAAACAAGATGAG-3'; -62kb, 5'-ACACATTTGTGCCAGTGCAG-3').

**Co-IP and western blotting analysis.** HEK293 cells were seeded for 1 day, transfected with the expression plasmids expressing wild-type, mutant AR and FoxA1 using Lipofectamine2000 (Invitrogen) and then changed to hormone-depleted, phenol-free DMEM medium. One day after plasmid transfection, cells were treated with 100 nM DHT for another day. Cells were washed by cold PBS twice and treated with 1 ml of lysis buffer (50 mM Tris pH 8.0, 150 mM NaCl, 1% NP-40) supplemented with a cocktail of proteinase inhibitors (Sigma) for 5 min at 4 °C. Lysed cells were collected, rotated for 1 h at 4 °C and cell debris removed by centrifugation at 18,000g for 30 min in a cold room. The supernatant was incubated with anti-AR, anti-FoxA1 or immunoglobulin-G overnight at 4 °C followed by the addition of 50 µl of 50% protein G beads to each tube. After rotating for another 2 h at 4 °C, the beads were washed five times with the lysis buffer, twice with cold PBS, and boiled for 6 min in 40 µl of 2 × SDS loading buffer. Western blotting analysis was performed with anti-AR or anti-FoxA1.

**Luciferase reporter assay.** PC3-AR and HEK293 cells were seeded into 24-well plates in hormone-depleted and phenol-free RPMI 1640 medium and DMEM 1 day before transfection. Transfection was performed according to the manufacturers' recommendations (DOTAP Liposomal Transfection Reagent from Roche or Lipofectamin 2000 from Invitrogen). One day after transfection, these cells were treated with DHT for an additional day. After washing with cold PBS twice, cells were treated with the lysis buffer (Promega) and the Luciferase signal was recorded.

**Cell proliferation assay.** The assay was based on the published protocol<sup>32</sup>. Briefly, LNCaP cells were transfected with control siRNA and FoxA1 siRNA (sequences listed in Methods Summary) and cultured in hormone-depleted medium for 3 days. The cells were treated with different amount of DHT for another day. After the treatment, cells were washed by PBS, fixed by 70% EtOH and stored at -20 °C for at least 2 h. Before analysis, cells were washed with cold PBS, resuspended at the propidium-iodide/Triton X-100 staining solution and incubated at 37 °C for 15 min. After removing cell clumps, stained cells were sorted on a Beckman FAScan, and the percentage of S-phase cells was calculated.

**ChIP and ChIP-seq analyses.** ChIP was as previously described<sup>29</sup>. Briefly, approximately 10<sup>7</sup> treated cells were crosslinked with 1% formaldehyde at room temperature for 15 min. After sonication, the soluble chromatin was incubated with 1–5 µg of antibody. Specific immunocomplexes were precipitated with Protein A/G beads (Sigma-Aldrich). Complexes were washed, DNA extracted and purified by QIAquick Spin columns (Qiagen). Extracted DNA (1 µl from 60 µl) was used for qPCR with the specific PCR primers listed in Supplementary

Fig. 24, each of which was designed surrounding a specific region of 150–250 base pairs (bp) on target DNA. PCR products were detected with SYBR Green on a MX3000P System (Stratagene) and the percentage of immunoprecipitated chromatin was calculated from  $\Delta C_t$  relative to immunoglobulin-G control after normalizing against input chromatin. For ChIP-seq, extracted DNA was ligated to specific adaptors followed by deep sequencing in the Illumina GAII system according to the manufacturer's instructions. The first 25 bp for each sequence tag returned by the Illumina Pipeline was aligned to the hg18 assembly (National Center for Biotechnology Information, build 36.1) using Bowtie, allowing up to two mismatches. Only tags uniquely mapped to the genome were used for further analysis. The data were visualized by preparing custom tracks for the University of California, Santa Cruz, genome browser using HOMER<sup>33</sup> (<http://biowhat.ucsd.edu/homer>). The total number of mappable reads was normalized to 10<sup>7</sup> for each experiment presented in this study. ChIP-seq at nucleosome resolution was performed as previously reported<sup>34</sup>. A summary of ChIP experiments is provided in Supplementary Fig. 25.

**Identification of ChIP-seq peaks.** The identification of ChIP-seq peaks (bound regions) was performed using HOMER (<http://biowhat.ucsd.edu/homer>). For transcription factors, peaks were identified by searching locations of high read density using a 200-bp sliding window. Regions of maximal density exceeding a given threshold were called as peaks, and we required adjacent peaks to be at least 500 bp away to avoid redundant detection. Only one tag from each unique position was considered to avoid clonal artefacts from the sequencing. The threshold for the number of tags that determined a valid peak was selected at a false discovery rate of 0.001 determined by peak finding using randomized tag positions in a genome with an effective size of  $2 \times 10^9$  bp. We also required peaks to have at least fourfold more tags (normalized to total count) than input control samples. In addition, we required fourfold more tags relative to the local background region (10 kb) to avoid identifying regions with genomic duplications or non-localized binding.

The peak finding procedure was modified to identify regions harbouring specific histone modifications, as these experiments tend to yield broad areas of enrichment over several hundreds or thousands of base pairs. Seed regions were initially found using a peak size of 500 bp at the false discovery rate of 0.001 to identify enriched loci. Enriched loci found within 1 kb of one another were then merged to yield variable-length regions. Transcription factor peaks and histone modification regions were associated with gene products by identifying the nearest RefSeq TSS. Annotated positions for promoters, exons, introns and other features were based on RefSeq transcripts and repeat annotations from University of California, Santa Cruz. Peaks from separate experiments were considered equivalent/co-bound if their peak centres were located within 200 bp of each other. Read density heat maps were created by first using HOMER to generate read densities and then visualized using Java TreeView (<http://jtreeview.sourceforge.net>).

**HOMER for de novo motif discovery and known motif enrichment.** Motif discovery was performed using a comparative algorithm similar to those previously described<sup>35</sup>. An in-depth description will be published elsewhere (Benner *et al.*, in preparation). Motif finding for transcription factors was performed on sequence from  $\pm 100$  bp relative to the peak centre, whereas motif finding for histone modification regions was performed on sequence from  $\pm 500$  bp relative to the region centre. Briefly, sequences were divided into target and background sets for each application of the algorithm. Background sequences were then selectively weighted to equalize the distributions of G + C content in target and background sequences to avoid comparing sequences of different general sequence content. Motifs of length 8, 10, 12, 14, 16 and 18 bp were identified separately by first exhaustively screening all oligonucleotides for enrichment in the target set compared with the background set using the cumulative hypergeometric distribution to score enrichment. Up to three mismatches were allowed in each oligonucleotide sequence to increase the sensitivity of the method. The top 200 oligonucleotides of each length with the lowest *P* values were then converted into probability matrices and heuristically optimized to maximize hypergeometric enrichment of each motif in the given data set. As optimized motifs were found they were removed from the data set to facilitate the identification of additional motifs in subsequent rounds. HOMER also screens the enrichment of known motifs previously identified through the analysis of published ChIP-ChIP and ChIP-Seq data sets by calculating the known motifs' hypergeometric enrichment in the same set of G + C normalized sequences used for *de novo* analysis. Sequence logos were generated using WebLOGO (<http://weblogo.berkeley.edu>). Motif enrichment heatmaps and dendrograms were created by clustering hypergeometric log *P* values using Cluster (<http://bonsai.ims.u-tokyo.ac.jp/~mdehoo/software/cluster/software.htm#ctv>) and Java TreeView (<http://jtreeview.sourceforge.net>).

**GRO-seq.** Global run-on<sup>6</sup> and library preparation for sequencing<sup>30</sup> were done as described. Briefly, four 10-cm plates of confluent LNCaP cells per treatment were washed three times with cold PBS buffer. Cells were then swelled in swelling buffer (10 mM Tris pH 7.5, 2 mM MgCl<sub>2</sub>, 3 mM CaCl<sub>2</sub>) for 5 min on ice. Harvested cells



were re-suspended in 1 ml of the lysis buffer (swelling buffer with 0.5% IGEPAL and 10% glycerol) with gentle vortex and brought to 10 ml with the same buffer for extraction of nuclei. Nuclei were washed with 10 ml of lysis buffer and re-suspended in 1 ml of freezing buffer (50 mM Tris pH 8.3, 40% glycerol, 5 mM MgCl<sub>2</sub>, 0.1 mM EDTA), pelleted down again and finally re-suspended in 100 µl of freezing buffer.

For run-on assay, re-suspended nuclei were mixed with an equal volume of reaction buffer (10 mM Tris pH 8.0, 5 mM MgCl<sub>2</sub>, 1 mM DTT, 300 mM KCl, 20 units of SUPERase-In, 1% Sarkosyl, 500 µM ATP, GTP and Br-UTP, 2 µM CTP) and incubated for 5 min at 30 °C. Nuclei RNA were extracted with TRIzol LS reagent (Invitrogen) according to the manufacturer's instructions. RNA was then re-suspended in 20 µl of DEPC-water and subjected to base hydrolysis by addition of 5 µl of 1 M NaOH and incubated on ice for 40 min. Then, 25 µl of 1 M Tris pH 6.8 was added to neutralize the reaction. RNA was purified through a p-30 RNase-free spin column (BioRad), according to the manufacturer's instructions and treated with 6.7 µl of DNase buffer and 10 µl of RQ1 RNase-free DNase (Promega), and purified again through a p-30 column. A volume of 8.5 µl 10 × antartctic phosphatase buffer, 1 µl of SUPERase-In and 5 µl of antartctic phosphatase was added to the run-on RNA and treated for 1 h at 37 °C. Before proceeding to immunopurification, RNA was heated to 65 °C for 5 min and kept on ice.

Anti-BrdU argarose beads (Santa Cruz Biotech) were blocked in blocking buffer (0.5 × SSPE, 1 mM EDTA, 0.05% Tween-20, 0.1% PVP, and 1 mg ml<sup>-1</sup> BSA) for 1 h at 4 °C. Heated run-on RNA (~85 µl) was added to 60-µl beads in 500 µl binding buffer (0.5 × SSPE, 1 mM EDTA, 0.05% Tween-20) and allowed to bind for 1 h at 4 °C with rotation. After binding, beads were washed once in low salt buffer (0.2 × SSPE, 1 mM EDTA, 0.05% Tween-20), twice in high salt buffer (0.5 × SSPE, 1 mM EDTA, 0.05% Tween-20, 150 mM NaCl) and twice in TET buffer (TE pH 7.4, 0.05% Tween-20). BrU-incorporated RNA was eluted with 4 × 125 µl elution buffer (20 mM DTT, 300 mM NaCl, 5 mM Tris pH 7.5, 1 mM EDTA and 0.1% SDS). RNA was then extracted with acidic phenol/chloroform once, chloroform once and precipitated with ethanol overnight. The precipitated RNA was re-suspended in 50 µl reaction (45 µl of DEPC water, 5.2 µl of T4 PNK buffer, 1 µl of SUPERase-In and 1 µl of T4 PNK (NEB)) and incubated at 37 °C for 1 h. The RNA was extracted and precipitated again as above.

Complementary DNA (cDNA) synthesis was performed basically as described<sup>30</sup> with minor modifications. First, RNA fragments were subjected to poly-A tailing reaction in 8.0 µl volume containing 0.8 µl poly-A polymerase buffer, 1 µl 1 mM ATP, 0.5 µl SUPERase-In and 0.75 µl poly-A polymerase (NEB). The reaction was performed for 30 min at 37 °C. Subsequently, reverse transcription was performed using oNTI223 primer (5'-pGATCGTCCGACTGTAGAACTCT;CAAGCAGAAGACGGCATACGATTTTCTTTTCTTTTCTTTTCTVN-3') where the p indicates 5' phosphorylation, 'i' indicates the abasic dSpacer furan and VN indicates degenerate nucleotides.

Tailed RNA (8.0 µl) was mixed with 1 µl dNTP (10 mM each) and 2.5 µl 12.5 µM oNTI223, heated for 3 min at 75 °C and chilled briefly on ice. Then, 0.5 µl SUPERase-In, 3 µl 0.1 M DTT, 2 µl 25 mM MgCl<sub>2</sub>, 2 µl 10 × reverse transcription buffer and 1 µl superscript III reverse transcriptase (Invitrogen) was added to the tube. The tube was incubated for 30 min at 48 °C. After that, 4 µl of Exonuclease I (Fermentas) was added into the reaction and the tube was incubated for 1 h at 37 °C to eliminate extra oNTI223. Then RNA was eliminated by adding 1.8 µl 1 M NaOH and incubated for 20 min at 98 °C. The reaction was neutralized with 1.8 µl of 1 M HCl. After running on a 10% polyacrylamide TBE-urea gel, the extended first-strand cDNA product was excised and recovered by soaking the grinded gel in DNA gel elution buffer (TE with 0.1% Tween-20 and 150 mM NaCl) overnight and then precipitated with ethanol.

Circularization of first-strand cDNA was performed by re-suspending cDNA in 9.5 µl reaction solution (7.5 µl water, 1 µl CircLigase buffer, 0.5 µl 1 mM ATP and 0.5 µl 50 mM MnCl<sub>2</sub>) and then adding 0.5 µl CircLigase (Epicentre). The reaction went for 1 h at 60 °C and then was heat-inactivated for 20 min at 80 °C. Circularized single-stranded DNA (ssDNA) was relinearized by adding 3.8 µl of 4 × relinearization supplement (100 mM KCl, 2 mM DTT) followed by 1.5 µl of *ApeI* (15u, NEB). The reaction was incubated for 1 h at 37 °C. Relinearized ssDNA was separated in a 10% polyacrylamide TBE-urea gel (Invitrogen) as described above. The relinearized product band was excised (~120–300 bp) and the DNA was recovered as described above.

The ssDNA template was amplified by PCR using the Phusion High-Fidelity enzyme (NEB) according to the manufacturer's instructions. The oligonucleotide primers oNTI200 (5'-CAAGCAGAAGACGGCAT-3') and oNTI201 (5'-AATGATACGGCGACACCGACAGGTTACAGATTCTACAGTCCGACG-3') were used to generate DNA for sequencing. PCR was performed with an initial 30-s denaturation at 98 °C, followed by 13 cycles of 10-s denaturation at 98 °C, 15-s annealing at 60 °C and 15-s extension at 72 °C. The PCR product was run on a non-denaturing 8% polyacrylamide TBE gel and recovered as mentioned before.

DNA was then sequenced on the Illumina Genome Analyser II according to the manufacturer's instructions with small RNA sequencing primer 5'-CGACAGGTTTCAGAGTTCTACAGTCCGACGATC-3'.

**De novo identification of GRO-Seq transcripts.** To identify transcription units in an unbiased manner, GRO-Seq read densities were analysed to classify genomic regions into contiguous transcripts using HOMER. GRO-Seq read densities were initially normalized using the GC content of individual reads to remove any systematic bias introduced by overall GC content variation between read libraries. To maximize read depth for transcript identification, all GRO-Seq libraries were merged to perform the initial transcript discovery, and later considered separately to identify regulated transcripts. For each strand of each chromosome, GRO-Seq read densities were calculated using a sliding window of 250 bp. Regions for which GRO-Seq reads could not be uniquely mapped (that is, repeats) were first identified and then read densities in these regions were estimated using upstream read densities from mappable regions to avoid ending predicted transcripts prematurely. Transcript initiation sites were identified as regions where the GRO-Seq read density increased threefold relative to the previous 1 kb region. Transcript termination sites were defined by either a reduction in reads below 10% of the start of the transcript or when another transcript's start site occurred on the same strand. Single spikes in read density covering a span less than 250 bp were considered artefacts and discarded. Identified transcripts were strand-specifically compared with RefSeq transcripts by looking for overlap in the transcribed region. Transcripts were defined as putative eRNAs if their TSS was located distal to RefSeq TSS (>3 kb) and were associated with H3K4me1 regions. To identify differentially regulated transcripts, strand-specific read counts from each GRO-Seq experiment were determined for each transcript using HOMER<sup>33</sup>. EdgeR (<http://www.bioconductor.org/packages/release/bioc/html/edgeR.html>) was then used to calculate differential expression (>1.5-fold, <0.01 false discovery rate).

**Microarray and reverse-transcription qPCR analyses of gene expression.** Total RNA was isolated with the RNeasy Mini Kit (Qiagen) and treated by RNase-free DNase I. For PCR with reverse transcription, first-strand cDNA synthesis from total RNA was performed with the Superscript III cDNA Synthesis System (Invitrogen). Microarray analysis was performed on Human V2 Chips (Illumina). The published gene expression profiling data GDS2545 (refs 36, 37) and GDS1439 (ref. 38) were extracted from the National Center for Biotechnology Information, normalized and *P* values calculated by two-tailed *t*-test. For validation by PCR with reverse transcription, cDNA was analysed with SYBR Green (Stratagene) on the Mcx300P System (Stratagene). The relative messenger RNA level was calculated by comparing with non-treatment control, after normalization with *GAPDH* or *ACTB* messenger RNA. The primers for RT-qPCR (5'-3') were as follows: ACTB-5, CGTCCCAGT TGGTGACGATG; ACTB-3, GCCGTCTTCCCCTCCATC; GAPDH-5, GTTTTCTAGACGGCAGGTTCAGG; GAPDH-3, AACATCATCCCTGCCTCTACTGG; KLK3-5, TGTGTGCGCAAGTTCAAC; KLK3-3, GGTTCACCTGCCCATGAC; RASSF3-5, GACGCCGAGGACTTCTTCTT; RASSF3-3, TGCTGAGGTAACGTGGGGTTT; SOX9-5, GACTCGCCACACTCCTCCTC; SOX9-3, AAGTCGATAGGGGGCTGTCT; IL6R-5, GAGATTCTGCAAATGCGACA; IL6R-3, GTGGG GAGATGACAGGAACA; DNM2-5, TGTTTGCCACAGTGACCTC; DNM2-3, CCCAGACAGTCAAGCTCCT.

**Survival analysis.** Two independent sets of gene expression data were used to check the association between FoxA1 and clinical outcome of patients by Kaplan-Meier analysis. One data set came from 78 patients with prostate tumours (age <70)<sup>39</sup>, the other from 131 patients<sup>40</sup>. Significant association with outcome was determined by log-rank test for survival. Hazard ratios were calculated by the Cox proportional model. All statistics were analysed with the statistical software R (version 2.6.2), available from the R Project for Statistical Computing website (<http://www.r-project.org>). The cut-off was determined so that the log-rank *P* value was the smallest one in the cut-offs that went through the 5th–95th percentiles of signals.

- Hu, Q. *et al.* Enhancing nuclear receptor-induced transcription requires nuclear motor and LSD1-dependent gene networking in interchromatin granules. *Proc. Natl Acad. Sci. USA* **105**, 19199–19204 (2008).
- Holbro, T. *et al.* The ErbB2/ErbB3 heterodimer functions as an oncogenic unit: ErbB2 requires ErbB3 to derive breast tumor cell proliferation. *Proc. Natl Acad. Sci. USA* **100**, 8933–8938 (2003).
- Heinz, S. *et al.* Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol. Cell* **38**, 576–589 (2010).
- Barski, A. *et al.* High-resolution profiling of histone methylations in the human genome. *Cell* **129**, 823–837 (2007).
- Linhardt, C., Halperin, Y. & Shamir, R. Transcription factor and microRNA motif discovery: the Amadeus platform and a compendium of metazoan target sets. *Genome Res.* **18**, 1180–1189 (2008).
- Chandran, U. R. *et al.* Gene expression profiles of prostate cancer reveal involvement of multiple molecular pathways in the metastatic process. *BMC Cancer* **7**, 64 (2007).

37. Yu, Y. P. *et al.* Gene expression alterations in prostate cancer predicting tumor aggression and preceding development of malignancy. *J. Clin. Oncol.* **22**, 2790–2799 (2004).
38. Varambally, S. *et al.* Integrative genomic and proteomic analysis of prostate cancer reveals signatures of metastatic progression. *Cancer Cell* **8**, 393–406 (2005).
39. Glinsky, G. V., Glinskii, A. B., Stephenson, A. J., Hoffman, R. M. & Gerald, W. L. Gene expression profiling predicts clinical outcome of prostate cancer. *J. Clin. Invest.* **113**, 913–923 (2004).
40. Taylor, B. S. *et al.* Integrative genomic profiling of human prostate cancer. *Cancer Cell* **18**, 11–22 (2010).

# Converting nonsense codons into sense codons by targeted pseudouridylation

John Karijovich<sup>1</sup> & Yi-Tao Yu<sup>1</sup>

All three translation termination codons, or nonsense codons, contain a uridine residue at the first position of the codon<sup>1–3</sup>. Here, we demonstrate that pseudouridylation (conversion of uridine into pseudouridine ( $\Psi$ ), ref. 4) of nonsense codons suppresses translation termination both *in vitro* and *in vivo*. *In vivo* targeting of nonsense codons is accomplished by the expression of an H/ACA RNA capable of directing the isomerization of uridine to  $\Psi$  within the nonsense codon. Thus, targeted pseudouridylation represents a novel approach for promoting nonsense suppression *in vivo*. Remarkably, we also show that pseudouridylated nonsense codons code for amino acids with similar properties. Specifically,  $\Psi$ AA and  $\Psi$ AG code for serine and threonine, whereas  $\Psi$ GA codes for tyrosine and phenylalanine, thus suggesting a new mode of decoding. Our results also suggest that RNA modification, as a naturally occurring mechanism, may offer a new way to expand the genetic code.

$\Psi$ , the C5-glycoside isomer of uridine, has many structural and biochemical differences from uridine<sup>5</sup> (Supplementary Fig. 1). Thus, it is possible that replacement of the uridine within a nonsense codon with  $\Psi$  may affect translation termination. To investigate the possible effect of  $\Psi$  on translation termination, we developed an *in vitro* nonsense suppression assay (Fig. 1a). Briefly, we synthesized an artificial messenger RNA that encoded a 6 $\times$  histidine (6His) tag at the amino terminus and a Flag tag at the carboxy terminus. In between the 6His tag and Flag tag, a pseudouridylated nonsense codon ( $\Psi$ AA) was inserted (Fig. 1a). In addition, two control transcripts were created with the same sequence except that the  $\Psi$  of the nonsense codon was either changed to uridine (U), thus forming an authentic nonsense codon (UAA), or substituted with a cytidine (C), thus eliminating the nonsense codon (CAA) (Fig. 1a). Anti-6His immunoblot analysis indicated that all three RNAs were equally translated in rabbit reticulocyte lysate (Fig. 1b, top panel). Remarkably, however, according to the anti-Flag blot, the presence of a  $\Psi$  within the termination codon resulted in robust nonsense suppression (Fig. 1b, lower panel). Specifically, the  $\Psi$ AA-containing transcript produced a strong Flag signal which is almost comparable to that produced by the CAA-containing transcript (Fig. 1b). In contrast, only a background level of Flag was produced when the UAA-containing transcript was used (Fig. 1b). Our results thus indicate that presence of  $\Psi$  in a termination codon effectively suppresses translation termination *in vitro*.

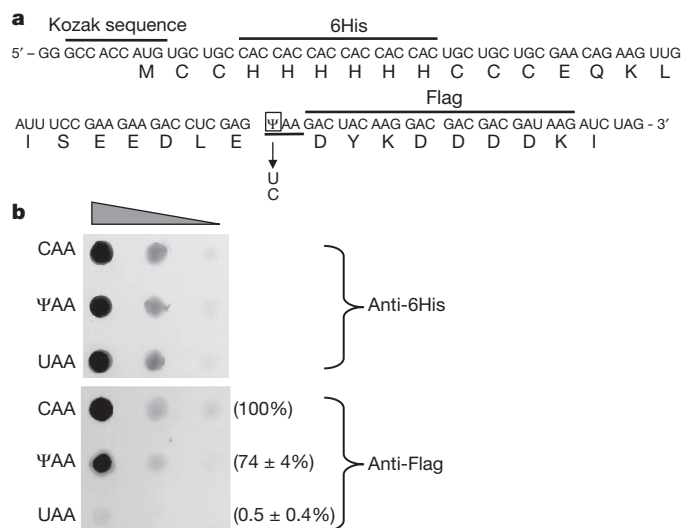
The *in vitro* results prompted us to investigate whether the pseudouridylation of a termination codon would elicit nonsense suppression *in vivo*. Taking advantage of the *CUP1* reporter system<sup>6</sup>, we introduced a premature termination codon (PTC) at the second codon of the *CUP1* gene, thus creating a new *CUP1* reporter gene (termed *cup1-PTC*). Cup1p is a copper chelating protein that mediates resistance to copper sulphate ( $\text{CuSO}_4$ )<sup>7</sup>. Thus, upon transformation of the *cup1-PTC* plasmid (pcup1-PTC) into a *Saccharomyces cerevisiae cup1-Δ* strain, one should be able to measure nonsense suppression by plating the cells on selective medium containing  $\text{CuSO}_4$  (Fig. 2A).

To direct site-specific  $\Psi$  formation *in vivo*, we took advantage of the H/ACA ribonucleoprotein (RNP) family. H/ACA RNPs are primarily responsible for the post-transcriptional isomerization of uridine to  $\Psi$

within RNA (Supplementary Fig. 2)<sup>8,9</sup>. To target the PTC within *cup1-PTC* we derived an H/ACA RNA from *SNR81*, a naturally occurring yeast H/ACA RNA. The newly derived H/ACA RNA, *snR81-1C*, contained a guide sequence capable of targeting the PTC within *cup1-PTC*. In addition, we also constructed a control H/ACA RNA, *snR81-Random*, which contained a random guide sequence.

To ensure that *cup1-PTC* was pseudouridylated in response to expression of *snR81-1C*, we measured  $\Psi$  formation within the PTC both *in vitro* and *in vivo*. To analyse  $\Psi$  formation *in vitro*, we monitored, by thin layer chromatography (TLC),  $\Psi$  formation on a 39-nucleotide fragment of RNA corresponding to the region of *cup1-PTC* containing the PTC (Fig. 2B). Incubation of the transcript in extracts prepared from cells expressing *snR81-1C* resulted in the formation of  $\Psi$  (Fig. 2B, lane 5), whereas extracts containing an empty vector or *snR81-Random* did not result in the formation of  $\Psi$  (lanes 3 and 4). These results indicate that *snR81-1C* is capable of directing pseudouridylation within the RNA fragment, most probably at the target site, that is, the uridine of the PTC.

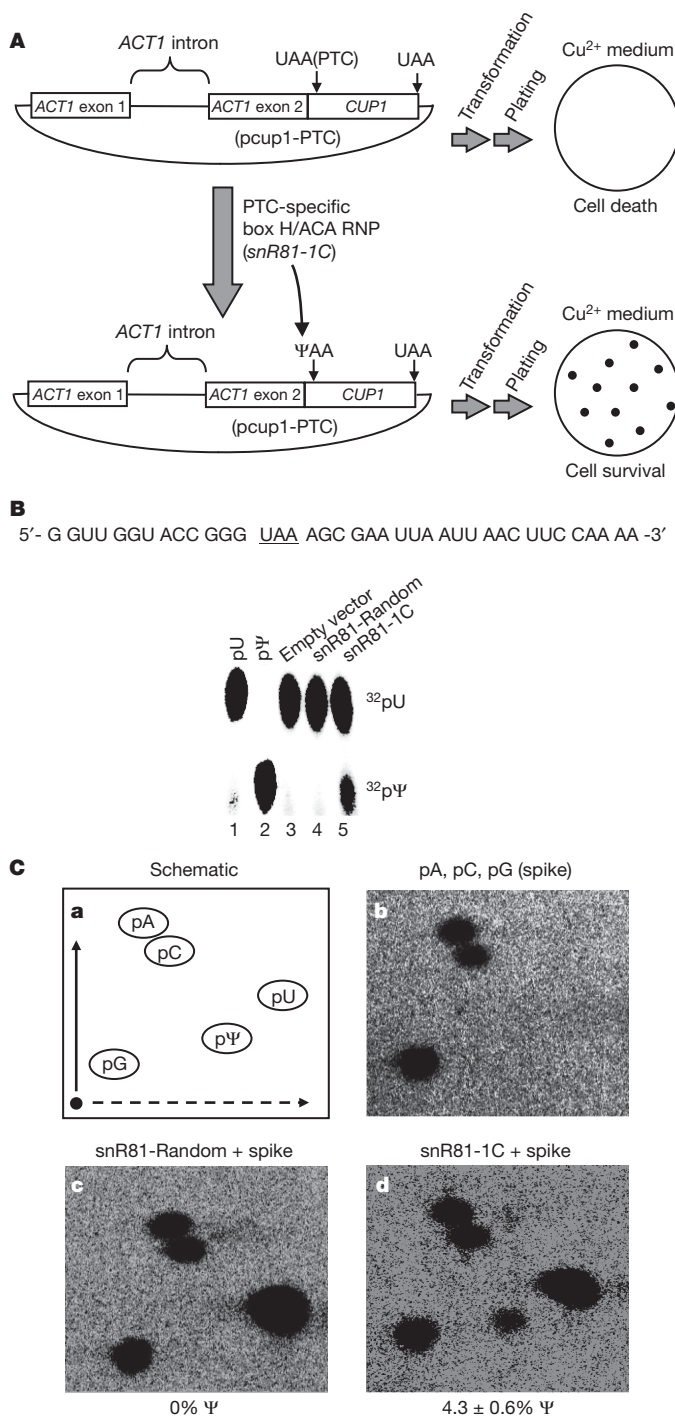
To determine the pseudouridylation status of *cup1-PTC* *in vivo*, we analysed the PTC of cellularly derived *cup1-PTC* mRNA using a site-specific and quantitative pseudouridylation assay, namely site-specific cleavage and radiolabelling followed by nuclease digestion and



**Figure 1 | Pseudouridylation of a termination codon promotes nonsense suppression *in vitro*.** **a**, Nucleotide sequence of the *in vitro* transcription product and its translated sequence are shown. Positions of the Kozak sequence, as well as epitopes (6His and Flag) within the nucleotide and protein sequences are labelled. The pseudouridylated nonsense codon is indicated. Changes of  $\Psi$  to U and  $\Psi$  to C are also indicated. **b**, Anti-6His and anti-Flag immunoblot analysis of the *in vitro* translation lysate following translation of an RNA lacking a termination codon (CAA), an RNA containing a pseudouridylated termination codon ( $\Psi$ AA), or an RNA containing an authentic termination codon (UAA). Relative efficiency of read-through (anti-Flag/anti-6His) was calculated and indicated in parentheses (the control, CAA, is set to 100%). Error is given as the standard deviation of three independent experiments.

<sup>1</sup>Department of Biochemistry and Biophysics, University of Rochester Medical Center, 601 Elmwood Avenue, Rochester, New York 14642, USA.





**Figure 2 | Quantification of *cup1-PTC* pseudouridylation.** **A**, Schematic of the *in vivo* nonsense suppression assay. **B**, *In vitro* pseudouridylation assay by thin layer chromatography. 5'  $^{32}\text{P}$ -radiolabelled uridylylated (pU) and pseudouridylylated (pΨ) markers were run in parallel. The substrate—[ $\alpha$ - $^{32}\text{P}$ ]UTP uniformly labelled RNA fragment—is shown. **C**, Quantification of *cup1-PTC* pseudouridylation *in vivo*. The percentage of pseudouridylation was calculated ( $\text{p}\Psi/(\text{p}\Psi + \text{pU})$ ). **a**, Schematic; **b**, Spike; **c**, *snR81-Random*; **d**, *snR81-1C*. Adenosine 5'-monophosphate (pA), cytidine 5'-monophosphate (pC), guanosine 5'-monophosphate (pG), uridine 5'-monophosphate (pU), and pseudouridine 5'-monophosphate (pΨ), are indicated. \*, origin. Error is given as the standard deviation of three independent experiments.

two-dimensional TLC (2D-TLC)<sup>10</sup>. To help locate the uridine and Ψ spots on the TLC plate, we spiked each reaction with  $^{32}\text{P}$ -radiolabelled adenosine 5'-monophosphate (pA), cytidine 5'-monophosphate (pC) and guanosine 5'-monophosphate (pG) (Fig. 2C, panel b). Consistent

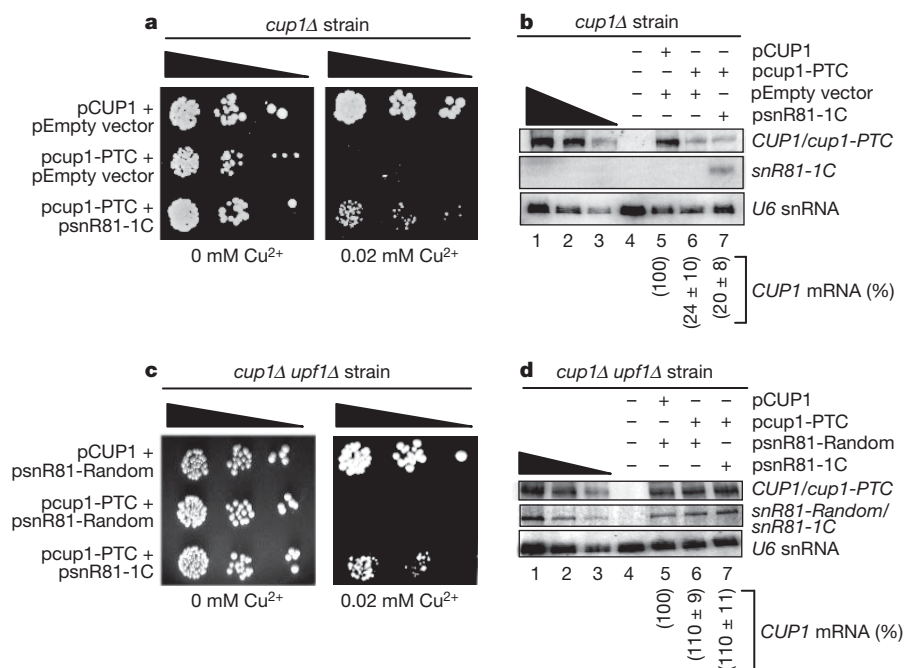
with the results of our *in vitro* analysis (Fig. 2B), only *CUP1* mRNA isolated from cells expressing *snR81-1C* produced a spot corresponding to Ψ (Fig. 2C, compare panel c with panel d). Quantification indicated that approximately 5% of the *cup1-PTC* transcript was pseudouridylylated. Thus, our *in vivo* pseudouridylation results would predict that, upon expression of *snR81-1C*, a functional Cup1p (although in small amount) would be translated from the *cup1-PTC* mRNA.

Figure 3a shows the results of the *in vivo* nonsense suppression assay (see Fig. 2a for illustration). As expected, when transformed with wild-type pCUP1, *cup1Δ* cells grew healthily on media containing 0 mM or 0.02 mM  $\text{CuSO}_4$  (top row). However, when transformed with pcup1-PTC, only cells co-transformed with psnR81-1C were able to survive on medium containing 0.02 mM  $\text{CuSO}_4$  (compare the middle row with the bottom row). Partial restoration of growth seems to be consistent with the previous quantitative analysis demonstrating a low level of pseudouridylation (~5%) at the PTC (Fig. 2c).

Because nonsense suppression could be achieved both at the level of translation (termination suppression) and the level of mRNA decay where PTC-containing mRNA is usually the target of NMD (nonsense-mediated decay), it remained possible that the suppression we observed in Fig. 3a was a result of NMD suppression rather than suppression of translation termination. To test this possibility, we measured the levels of *CUP1* mRNA using northern blot analysis (Fig. 3b). As expected, steady-state levels of *cup1-PTC* mRNA dropped significantly when compared with the level of wild-type *CUP1* mRNA (compare lanes 6 and 7 with lane 5). However, expression of the PTC-specific guide RNA (*snR81-1C*) had no effect on steady-state levels of *cup1-PTC* mRNA; nearly identical levels of *cup1-PTC* mRNA were detected in cells transformed with either psnR81-1C or with empty vector (compare lane 6 with lane 7). These results indicated that the observed suppression was a result of nonsense codon suppression rather than a result of suppressing NMD. To completely eliminate the potential complications of NMD, we deleted *UPF1* (also known as *NAM7*), a gene required for NMD<sup>11</sup>, and then repeated the nonsense suppression assay and northern blotting. As expected, deletion of *UPF1* resulted in the stabilization of *cup1-PTC* (Fig. 3d). Consistent with previously published results<sup>12,13</sup>, we also observed a small degree of nonsense suppression as evidenced by a low (but above background) level of growth on medium containing a low concentration (0.013 mM) of  $\text{CuSO}_4$  (Supplementary Fig. 3, compare row 5 with row 2). However, no growth was observed when  $\text{CuSO}_4$  concentration was raised to 0.02 mM (Fig. 3c, middle row). Apparently, the nonsense suppression phenotype conferred by deletion of *UPF1* is not sufficient to promote growth on 0.02 mM  $\text{CuSO}_4$ . To assess nonsense suppression induced by PTC pseudouridylation, we plated cells on media containing either 0.02 mM or 0.013 mM  $\text{CuSO}_4$  (Fig. 3c and Supplementary Fig. 3). Under both conditions, expression of *snR81-1C* provided a growth advantage; the level of growth rescue is similar to that observed when *UPF1* was intact (compare Fig. 3c with Fig. 3a, and Supplementary Fig. 3). These results further support the notion that expression of *snR81-1C* or pseudouridylation of the PTC resulted in suppression of translation termination rather than suppression of NMD. Given that the control, where *snR81-Random* was similarly expressed (Fig. 3d), showed no suppression (Fig. 3c), our results also indicate that the observed suppression of translation termination is guide-RNA-specific.

To determine further whether Ψ-mediated nonsense suppression can be generalized as well as which amino acids are incorporated at Ψ-substituted nonsense codons, we took advantage of a plasmid containing a C-terminally tagged *TRM4* gene (also known as *NCL1*), pTRM4-WT (Fig. 4a). Through site-directed mutagenesis the codon for phenylalanine at position 602 (F602) of the *TRM4* gene was changed to a nonsense codon (TAA, TAG or TGA), creating three variants of the plasmid (pTRM4-F602X(TAA), pTRM4-F602X(TAG), and pTRM4-F602X(TGA)) (Fig. 4a).

Figure 4b shows the western blot analysis of extracts prepared from wild-type cells expressing wild-type *TRM4* (nonsense-free) or

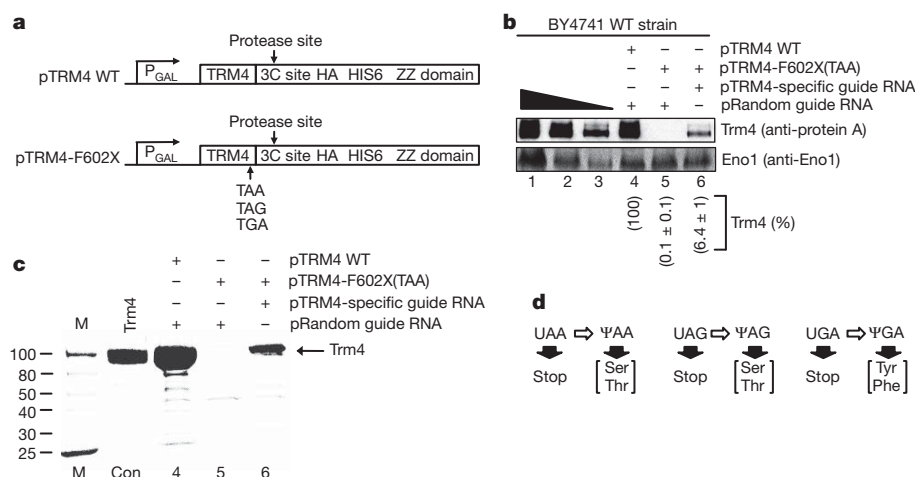


**Figure 3** | Expression of an H/ACA RNA targeting the PTC of *cup1-PTC* for pseudouridylation promotes nonsense suppression. **a**, pCUP1 or pcup1-PTC along with either an empty vector or psnR81-1C were transformed into a *cup1Δ* strain. Cell growth was assessed on solid synthetic medium (–Ura –Leu) containing either 0 mM or 0.02 mM CuSO<sub>4</sub>, as indicated. **b**, Northern blot analysis of RNA extracted from cells described in **a**. Normalized levels of *CUP1* mRNA (lane 5) and *cup1-PTC* mRNA (lanes 6 and 7) are indicated in parentheses under each lane. Error is given as the standard deviation of three independent experiments. **c**, *cup1Δ upf1Δ* strain was transformed with either pCUP1 or pcup1-PTC along with either psnR81-Random or psnR81-1C. Cell growth was assessed on solid synthetic medium (–His –Ura –Leu) with or without CuSO<sub>4</sub>, as indicated. **d**, Northern blot analysis of RNA extracted from cells described in **c**. Normalized levels of *CUP1* mRNA (lane 5) and *cup1-PTC* mRNA (lanes 6 and 7) are indicated in parentheses under each lane. Error is given as the standard deviation of three independent experiments.

*TRM4-F602X(TAA)*. When cells were transformed with the wild-type *TRM4* plasmid, a strong protein A-tag signal was detected (Fig. 4b, lane 4). However, when cells were transformed with pTRM4-F602X(TAA), a protein A signal was only detected when co-transformed with a *TRM4-F602X(TAA)*-specific guide RNA, indicating nonsense suppression (Fig. 4b, compare lane 5 with lane 6).

Next, we carried out large-scale immunoprecipitations to purify full-length Trm4p produced as a consequence of Ψ-mediated nonsense suppression. Figure 4c shows an example [pTRM4-F602X(TAA)] of such experiments. The bands corresponding to full-length Trm4p produced as a consequence of Ψ-mediated nonsense suppression were excised and sequenced by liquid chromatography–tandem mass spectrometry (LC-MS/MS; Fig. 4c and Supplementary Figs 4–7). Remarkably, LC-MS/MS analysis indicated that pseudouridylated

UAA and UAG (ΨAA and ΨAG) both directed the incorporation of either serine or threonine (Fig. 4d and Supplementary Figs 4–6). Taking into account that the third base of a codon is usually non-specific (the wobble base), it makes sense that both ΨAA and ΨAG code for the same amino acids. With respect to targeted pseudouridylation of UGA (ΨGA), it directed the incorporation of tyrosine and phenylalanine (Fig. 4d and Supplementary Fig. 7). As all three termination codons directed the incorporation of two amino acids, we quantified their frequency of incorporation (Supplementary Fig. 8). Interestingly, although ΨAA and ΨAG both code for serine and threonine, serine is predominantly incorporated at ΨAG, whereas serine and threonine are incorporated at a roughly similar frequency for ΨAA. Furthermore, ΨGA primarily directs the incorporation of tyrosine.



**Figure 4** | Generalization of Ψ-mediated nonsense suppression and determination of amino acids coded for by pseudouridylated nonsense codons. **a**, Schematic representation of the constructs used for protein purification (also see text). **b**, Western blot analysis was carried out using extracts prepared from wild-type cells transformed with either pTRM4 wild type (WT) and a plasmid containing a random guide RNA gene (pRandom guide RNA) (lane 4), pTRM4-F602X(TAA) and pRandom guide RNA (lane 5), or pTRM4-F602X(TAA) and a plasmid containing a guide RNA gene that targets the nonsense codon (UAA 602) of TRM4-F602X(TAA) (lane 6).

Enolase (Eno1) was probed as a loading control. The normalized levels of Trm4p are indicated in parentheses under each lane. Error is represented as the standard deviation from three independent experiments. **c**, Cell cultures described in **b** were scaled up, and Trm4 proteins were purified and analysed on a SDS-PAGE gel (stained with Coomassie blue); lanes correspond to those in **b**. In the control lane (Con), a known amount (6 μg) of purified Trm4p was loaded. M, molecular weight marker. **d**, Identification of amino acids incorporated at Ψ-containing termination codons (also see Supplementary Figs 4–8).

Interestingly, however, given that the anticodons of the transfer RNA<sup>Ser</sup> and tRNA<sup>Thr</sup> families do not look alike (Supplementary Fig. 9), our experimental data raise an important question: how is the same pseudouridylated nonsense codon (for example, ΨAA or ΨAG) recognized by the different families of tRNA? Although it is possible that the presence of Ψ in mRNA–tRNA duplexes acts to stabilize interactions between the mRNA and near- or non-cognate tRNAs<sup>14,15</sup>, an alternative explanation is that the presence of Ψ within the A-site may disorder the local RNA (or ribosome) structure, somehow allowing for the binding or accommodation of near- or non-cognate tRNAs, possibly through altering the hydration state of the nonsense codon<sup>16</sup>. It is also possible that unique RNA modifications in the anti-codon loop of tRNA<sup>Ser</sup>, tRNA<sup>Thr</sup>, tRNA<sup>Phe</sup> or tRNA<sup>Tyr</sup> contribute to the recognition of pseudouridylated nonsense codons, thus allowing them to be decoded. In fact, modifications within the anticodon loop of tRNA have previously been demonstrated to impact recoding<sup>17</sup>. Perhaps more interestingly, it has not escaped our attention that the amino acids incorporated at each termination codon are biochemically and structurally similar. Specifically, serine and threonine, which are coded for by ΨAA and ΨAG, are both hydroxylated short-chain amino acids. Likewise, tyrosine and phenylalanine, which are coded for by ΨGA, both contain an aromatic ring. Although the decoding centre is ~75 Å away from the peptidyl transferase centre<sup>18</sup>, whether there is a role for the amino acid in the decoding of Ψ-containing termination codons is an interesting idea that requires further analysis. If true, such a mechanism would represent a completely new mode of decoding. It is interesting to note that frameshifting at sense codons also shows a strong preference for using tRNA<sup>Ser</sup> and tRNA<sup>Thr</sup> (ref. 19). Although detailed mechanisms are still unclear, the similarities in using similar polar amino acids (serine and threonine) in frameshifting and in decoding of pseudouridylated nonsense codons certainly deserve further attention.

Our data demonstrate that artificial H/ACA guide RNAs are able to direct the pseudouridylation of nonsense codons of mRNA, thus leading to nonsense suppression. It should be noted that artificial guide RNAs may have an unintended target(s), thus raising concerns about substrate specificity. We did, however, realize this concern when designing sense-to-nonsense mutations and their corresponding guide RNAs, and purposely avoided the sites and their guide sequences that could target other endogenous mRNAs. In fact, BLAST search against the yeast genome did not generate any other potential targets that appear to be suitable substrates for our artificial H/ACA RNAs. Thus, it is unlikely that the observed effects are due to the nonspecific effect of modifications of unintended off-targets.

Our RNA-guided modification strategy is of significant clinical interest, given the current estimates that approximately 33% of genetic diseases can be attributed to the presence of a PTC<sup>20</sup>. On the other hand, because the artificial guide RNAs are derived from naturally occurring H/ACA RNAs (only the short guide sequence is changed), we predict that the nonsense codons of some mRNAs are naturally pseudouridylated by endogenous H/ACA RNAs as long as the guide sequence matches the target. Indeed, using computational algorithms to predict nonsense codons that may be natural targets of the endogenous H/ACA RNP machinery yields a number of potential candidates (Supplementary Fig. 10). In addition, our lab has recently demonstrated that an exact match between the H/ACA RNA guide sequence and the target sequence is not necessary for efficient modification under certain conditions. In fact, the mismatches are required for inducible pseudouridylation in response to cell stress<sup>21</sup>. Thus, there are probably a large number of pseudouridylation targets in mRNAs. Whereas some of these potential targets are nonsense codons, a majority of them are expected to be sense codons. Given our surprising discovery that pseudouridylation of nonsense codons converts them into sense codons, it is not impossible that pseudouridylation of sense codons will alter their decoding, making mRNA pseudouridylation a novel mechanism of RNA editing. If this is true, the genetic code would expand considerably. We predict

that targeted pseudouridylation of mRNA is a yet-to-be appreciated mechanism of generating protein diversity.

## METHODS SUMMARY

Relevant properties of strains and growth conditions are described in the text. Strain construction and additional growth conditions are described in the Methods. Standard procedures were used for all protein and RNA analyses and are described in the Methods. Mass spectrometry was performed at the University of Rochester Medical Center Proteomics Core and is described in the Methods.

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

Received 26 January; accepted 28 April 2011.

- Brenner, S., Barnett, L., Katz, E. R. & Crick, F. H. UGA: a third nonsense triplet in the genetic code. *Nature* **213**, 449–450 (1967).
- Brenner, S., Stretton, A. O. & Kaplan, S. Genetic code: the 'nonsense' triplets for chain termination and their suppression. *Nature* **206**, 994–998 (1965).
- Weigert, M. G. & Garen, A. Base composition of nonsense codons in *E. coli*. Evidence from amino-acid substitutions at a tryptophan site in alkaline phosphatase. *Nature* **206**, 992–994 (1965).
- Cohn, W. E. 5-Ribosyl uracil, a carbon-carbon ribofuranosyl nucleoside in ribonucleic acids. *Biochim. Biophys. Acta* **32**, 569–571 (1959).
- Charette, M. & Gray, M. W. Pseudouridine in RNA: what, where, how, and why. *IUBMB Life* **49**, 341051 (2000).
- Lesser, C. F. & Guthrie, C. Mutational analysis of pre-mRNA splicing in *Saccharomyces cerevisiae* using a sensitive new reporter gene, *CUP1*. *Genetics* **133**, 851–863 (1993).
- Hamer, D. H., Thiele, D. J. & Lemontt, J. E. Function and autoregulation of yeast copperthionein. *Science* **228**, 685–690 (1985).
- Ganot, P., Bortolin, M. L. & Kiss, T. Site-specific pseudouridine formation in preribosomal RNA is guided by small nucleolar RNAs. *Cell* **89**, 799–809 (1997).
- Ni, J., Tien, A. L. & Fournier, M. J. Small nucleolar RNAs direct site-specific synthesis of pseudouridine in ribosomal RNA. *Cell* **89**, 565–573 (1997).
- Zhao, X. & Yu, Y. T. Detection and quantitation of RNA base modifications. *RNA* **10**, 996–1002 (2004).
- Leeds, P., Peltz, S. W., Jacobson, A. & Culbertson, M. R. The product of the yeast *UPF1* gene is required for rapid turnover of mRNAs containing a premature translational termination codon. *Genes Dev.* **5**, 2303–2314 (1991).
- Leeds, P., Wood, J. M., Lee, B. S. & Culbertson, M. R. Gene products that promote mRNA turnover in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* **12**, 2165–2177 (1992).
- Wilusz, C. J., Wang, W. & Peltz, S. W. Curbing the nonsense: the activation and regulation of mRNA surveillance. *Genes Dev.* **15**, 1781–1785 (2001).
- Agris, P. F. The importance of being modified: roles of modified nucleosides and Mg<sup>2+</sup> in RNA structure and function. *Prog. Nucleic Acid Res. Mol. Biol.* **53**, 79–129 (1996).
- Davis, D. R. Stabilization of RNA stacking by pseudouridine. *Nucleic Acids Res.* **23**, 5020–5026 (1995).
- Auffinger, P. & Westhof, E. RNA hydration: three nanoseconds of multiple molecular dynamics simulations of the solvated tRNA<sup>ASP</sup> anticodon hairpin. *J. Mol. Biol.* **269**, 326–341 (1997).
- Agris, P. F. Decoding the genome: a modified view. *Nucleic Acids Res.* **32**, 223–238 (2004).
- Song, H. *et al.* The crystal structure of human eukaryotic release factor eRF1—mechanism of stop codon recognition and peptidyl-tRNA hydrolysis. *Cell* **100**, 311–321 (2000).
- Atkins, J. F., Gesteland, R. F., Reid, B. R. & Anderson, C. W. Normal tRNAs promote ribosomal frameshifting. *Cell* **18**, 1119–1131 (1979).
- Frischmeyer, P. A. & Dietz, H. C. Nonsense-mediated mRNA decay in health and disease. *Hum. Mol. Genet.* **8**, 1893–1900 (1999).
- Wu, G., Xiao, M., Yang, C. & Yu, Y. T. U2 snRNA is inducibly pseudouridylated at novel sites by Pus7p and snR81 RNP. *EMBO J.* **30**, 79–89 (2010).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank F. Hagen and the Proteomics Core at the University of Rochester for performing the mass spectrometry analysis. We also thank E. Phizicky and B. Grayhack for the wild-type TRM4 construct, C. Guthrie for the *cup1Δ* yeast strain, D. Mcpheeters for the wild-type CUP1 construct, and M. Dumont for the anti-Eno1p antibody. Lastly, we would like to thank members of the Yu laboratory, especially X. Zhao, for helpful discussions.

**Author Contributions** J.K. and Y.-T.Y. designed and interpreted the experiments. Mass spectrometry was performed at the Proteomics Core at the University of Rochester Medical Center. J.K. performed all other experiments.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at [www.nature.com/nature](http://www.nature.com/nature). Correspondence and requests for materials should be addressed to Y.-T.Y. ([yitao\\_yu@urmc.rochester.edu](mailto:yitao_yu@urmc.rochester.edu)).



## METHODS

**Yeast strains, transformation and growth assay.** The *cup1-Δ* yeast strain was kindly provided by C. Guthrie<sup>6</sup>. The *UPF1* locus was deleted from a *cup1-Δ* strain using a standard protocol as described previously<sup>22</sup>. For the analysis of CuSO<sub>4</sub> resistance the appropriate plasmids were transformed into either *cup1-Δ* or *cup1-Δ upf1-Δ* yeast strains as previously described<sup>22</sup>, except that after heat shock cells were precipitated and resuspended in 100 μl of water rather than YPD (yeast peptone dextrose). Single colonies were selected and grown to saturation in SGal (synthetic galactose) drop-out media, cells were diluted to an OD<sub>600</sub> = 0.001 and then a series of fivefold dilutions were spotted on to SGal drop-out media, with or without CuSO<sub>4</sub>. Growth phenotypes were assessed after cells were grown for 3–5 days at 30 °C.

**Plasmids.** The pCUP1 plasmid was a gift from D. Mcpheeters, and pTRM4 WT was a gift from E. Phizicky and B. Grayhack. pcup1-PTC and pTRM4 F602X variants were created by site-directed mutagenesis using Pfu polymerase (Stratagene) and the appropriate oligonucleotides and plasmids. Novel H/ACA RNA genes were constructed by PCR using four overlapping oligonucleotide primers and were either cloned into 2 μm *URA3* or 2 μm *LEU2* vector (both gifts from E. Phizicky) as BamHI/HindIII fragments<sup>23</sup>.

**In vitro transcription and translation.** To generate mRNA transcripts for *in vitro* translation, DNA templates were synthesized through PCR using two overlapping DNA oligonucleotides. The double-stranded DNA templates thus synthesized contained either a TAA nonsense codon or a CAA codon in the middle, flanked by a 6His-coding sequence near the 5' end and a Flag-coding sequence at the 3' end (Fig. 1b). For efficient *in vitro* translation, the templates also contained a Kozak sequence immediately upstream of the 6His coding sequence (Fig. 1b). In addition, a T7 promoter sequence was included at the 5' terminus. Following *in vitro* T7 transcription<sup>24,25</sup>, UAA- or CAA-containing mRNA transcripts were synthesized (see Fig. 1b). To create a similar mRNA, with the uridine of the nonsense codon (or the cytidine of the CAA codon) changed to Ψ, a two-piece splint ligation was employed<sup>24</sup>. The 5' RNA was *in-vitro*-synthesized through T7 transcription, ending with CUC at its 3' terminus (see Fig. 1b), and the 3' piece (5'-GAGΨAAGACUACAAGGACGACGACGACAAGAUCUAG-3') (see Fig. 1a) was chemically synthesized (Thermo Scientific). The 5' and 3' halves were ligated using a bridging oligonucleotide and T4 DNA ligase<sup>24</sup>. *In-vitro*-synthesized RNAs were gel-purified before being used in the *in vitro* translation reactions. *In vitro* translation reactions were carried out in 30 μl reactions of Red Nova Lysate as described by the supplier (Novagen). PCR of two overlapping oligodeoxynucleotides was also used to generate the template for *in vitro* transcription of the substrate used in the *in vitro* pseudouridylation assay (Fig. 2C).

**Northern blot analysis.** Total RNA was isolated from yeast using TRIzol essentially as described by the supplier (Invitrogen), except that cells were vortexed with acid-sterilized glass beads for 5 min. For northern blot analysis, 6 μg of total RNA was separated on 8% polyacrylamide–7 M urea gels and electrotransferred at 4 °C to Amersham Hybond-N<sup>+</sup> membranes in 0.5× TBE buffer for 16 h at 15 V. Hybridizations were performed essentially as described<sup>23</sup>.

**Protein purification and immunoblot analysis.** For the analysis of Trm4p protein sequence, BY4741 was transformed with the appropriate plasmids as described before<sup>22</sup>. Yeast whole-cell extracts and IgG Sepharose chromatography was performed as previously described<sup>26</sup>. For analysis of *in vitro* translation products, membranes were probed with either a monoclonal Flag antibody (M2; Sigma-Aldrich) or monoclonal His-probe (H-3; Santa Cruz Biotechnology). Goat

anti-mouse IgG (H+L)–alkaline phosphatase (AP) conjugate (Bio-Rad) was then used as a secondary antibody. Proteins were visualized using 1-Step NBT/BCIP (Pierce). For the analysis of Trm4p, yeast crude extracts were separated on 4–15% Tris-HCl Ready gels (BioRad). Proteins were then transferred to 0.1 μm nitrocellulose membranes (Whatman) and probed with either monoclonal Protein A (Sigma-Aldrich) or anti-Eno1p (a gift from M. Dumont). Goat anti-Mouse IgG (H+L)–AP conjugate (Bio-Rad) was used as a secondary antibody. Proteins were visualized using 1-Step NBT/BCIP (Pierce).

**Pseudouridylation assays.** *In vitro* pseudouridylation assays were performed using yeast whole-cell extract. Cells were grown to mid log phase and pelleted. Pellets were resuspended in 200 μl of extraction buffer containing 20 mM HEPES at pH 7.9, 0.42 M NaCl, 1.5 mM MgCl<sub>2</sub>, 0.2 mM EDTA, 0.5 mM DTT, 0.5 mM PMSF, and 25% glycerol. Sterile acid-washed glass beads (400 μl) were added to the cell suspension, and cells were subsequently homogenized through vigorous vortexing (5 × 30 s) at 4 °C. Following a 5-min centrifugation (14,000g, 4 °C), the supernatant was recovered, and used for the pseudouridylation assay. The substrate RNA was prepared by *in vitro* transcription in the presence of [ $\alpha$ -P<sup>32</sup>]UTP. The substrate was gel purified and incubated in the extracts for 2 h. The radio-labelled substrate was recovered and digested with nuclease P1 and analysed by one-dimensional-TLC as previously described<sup>10</sup>. Pseudouridylation of cellularly derived *cup1*-PTC RNA was analysed as previously described<sup>10</sup>, except that modifications were analysed by 2D-TLC<sup>27</sup>.

**Mass spectrometry.** Mass spectrometry was performed at the University of Rochester Proteomics Center. Coomassie-stained gel bands corresponding to full-length Trm4p were subjected to in gel trypsin digestion. An 80-min LC-MS/MS run was performed in-line with a Finnigan LTQ Ion Trap mass spectrometer (Thermo Scientific), using a flow rate of 250 μl min<sup>-1</sup>. The data collected from the LTQ runs was searched using MASCOT (Matrix Science), initially against the full *Saccharomyces* database, second against a custom database which included the wild type Trm4p sequence, as well as a Trm4p sequence with an "X" in the amino acid position that corresponds to the stop codon. Peptides identified by Mascot with an ion score of 15 or greater were inspected further for MS/MS fragmentation patterns that map through most of the peptide sequence, especially on and through the mutant amino acid position. Peptides with Expect values greater than 0.05 were not accepted. To allow for relative quantification, we repeated the LC-MS/MS experiments using dynamic inclusion for only the peptides of interest. Total spectral counts obtained by dynamic inclusion therefore represent the relative abundances of each respective peptide.

22. Ma, X. *et al.* Pseudouridylation of yeast U2 snRNA is catalyzed by either an RNA-guided or RNA-independent mechanism. *EMBO J.* **24**, 2403–2413 (2005).
23. Chernyak, I., Whipple, J. M., Kotelawala, L., Grayhack, E. J. & Phizicky, E. M. Degradation of several hypomodified mature tRNA species in *Saccharomyces cerevisiae* is mediated by Met22 and the 5'-3' exonucleases Rat1 and Xrn1. *Genes Dev.* **22**, 1369–1380 (2008).
24. Yu, Y. T. Site-specific 4-thiouridine incorporation into RNA molecules. *Methods Enzymol.* **318**, 71–88 (2000).
25. Zhao, X. & Yu, Y. T. Pseudouridines in and near the branch site recognition region of U2 snRNA are required for snRNP biogenesis and pre-mRNA splicing in *Xenopus oocytes*. *RNA* **10**, 681–690 (2004).
26. Gelperin, D. M. *et al.* Biochemical and genetic analysis of the yeast proteome with a movable ORF collection. *Genes Dev.* **19**, 2816–2826 (2005).
27. Zebajadian, Y., King, T., Fournier, M. J., Clarke, L. & Carbon, J. Point mutations in yeast *CBF5* can abolish *in vivo* pseudouridylation of rRNA. *Mol. Cell. Biol.* **19**, 7461–7472 (1999).

# CAREERS

**GERMANY** New collaborative-research centres seek junior scientists **p.409**

**ACADEMIA** Isolation and lack of confidence hold back female postdocs **p.409**

**NATUREJOBS** For the latest career listings and advice [www.naturejobs.com](http://www.naturejobs.com)

S. VIDLER/EURASIA PRESS/CORBIS



FIELDWORK

## Close quarters

*Cramped living conditions, unruly colleagues or crowded schedules can be challenging for the most intrepid scientist.*

BY LUCAS LAURSEN

In the scientists' lounge aboard the *BIO Hespérides* one evening last March, Jordi Dachs points at the schedule for the next day's oceanographic observations. The Spanish research vessel is chugging across the Indian Ocean at a speed of about ten knots. "The storm has put us seven hours behind," warns Dachs, an environmental chemist at the Institute of Environmental Assessment and Water Research in Barcelona, Spain, whose responsibilities as chief scientist on the ship include

planning researchers' time and instrument use. About two dozen scientists brace themselves against the rhythmic pitching of the vessel. "We might not lower the sampling rosette all the way on some days," says Dachs, "to save time."

His suggestion fills the room with tension. Lowering and retrieving the rosette can take many hours, but the water samples it retrieves from the ocean's depths — as much as 4 kilometres down — hold the biggest potential for new discoveries. Sampling excursions in the Indian Ocean's deep waters are relatively rare, making the samples particularly valuable.

"Isn't this a deep-sea campaign?" snaps Celia Marrasé, a plankton ecologist at the Institute of Marine Sciences in Barcelona. Marrasé studies how dissolved and particulate organic matter sequesters carbon in the sea: skipping the deepest waters would affect her results. After more heated discussion, Dachs agrees to consider saving a few minutes on each of the 12 remaining days at sea by cutting time for all the observations, not just those from the rosette. The disgruntled researchers scatter, but cannot go very far: their bunk rooms open directly onto the scientists' lounge.

In such a tight space, there is neither the room, the time nor the money to let differences escalate. "It's very expensive to send people to these places and you want to get a lot of return," says Albert Harrison, a retired psychologist from the University of California, Davis, who has studied the social factors that lead to success on space missions. For decades, NASA, the Russian Federal Space Agency and researchers in psychology and anthropology have examined how to achieve productivity in remote settings, gleaned lessons that are useful for any scientist who conducts fieldwork in close quarters.

Astronauts, polar biologists, desert geologists and ocean chemists all face similar challenges in their working conditions: relative isolation; overexposure to their colleagues; pressure to accomplish a lot of work in a short time with limited technical resources; and few options for escaping to family or friends outside work. If managed poorly, such circumstances can degrade a team's cohesion, health and productivity. Scientists' experiences, and the work of psychologists such as Harrison, suggest that researchers designing remote expeditions need to cultivate a sense of mission to guard against tough conditions, plan ahead with suitable schedules and equipment, practise working in field conditions and, above all, exercise patience when colleagues, instruments and experiences don't live up to their initial hopes.

### FORWARD PLANNING

In the early years of human space flight, mission planners had huge expectations for how the astronauts would spend their precious days in space. "Historically there was a tendency to over-programme," says Harrison. That was exacerbated by the enormous cost and limited flight time of each expedition. ►



► On a smaller scale, the same thing often happens in fieldwork on Earth. When Tim Wright, a geophysicist at the University of Leeds, UK, led his first field trip to the Afar region of Ethiopia, he had a long list of multi-disciplinary observations that he hoped to make, from collecting rocks to mapping fault structures. “We go in with these incredibly detailed plans: day one, go here; day two, go here,” says Wright. By the second day, “it’s usually ripped up and used to start the fire, because things take way longer than you anticipate”.

Trying too hard to achieve too much can harm the scientists and their data. Alberto Escribano, second in command on the *Hespérides*, is in the Spanish navy, and has first-hand experience of the consequences of teams pushing themselves too far. He explains that Spanish navy crews typically alternate 6-hour work and rest shifts on combat ships, but if they maintain this schedule for much longer than two weeks, the crew members become exhausted — and havoc can ensue. Sailors who are overtired don’t cope well with tasks that demand close attention, such as keeping the ship steady while instruments hang overboard. The same effect can wear down researchers, leading to squashed fingers, broken sampling bottles and fewer or lower-quality data. The solution for longer voyages, such as seven-month cruises on the *Hespérides*, is for each crew member to take shorter watches with longer recovery times.

“You want to avoid cumulative fatigue,” says Harrison. “Rest periods are not just perks — they help workers maintain their strength and high level of productivity.”

Conflicting research goals can overload an expedition’s agenda and put researchers under stress. The *Hespérides* expedition led by Dachs is just one leg of a Spanish National Research Council (CSIC) circumnavigation of the Indian Ocean that is trying to balance the research objectives of six disciplines, including biogeochemistry, microbiology and

environmental chemistry. The ship has to be stationary for researchers to use some instruments, such as the sampling rosette, but has to be moving for others, such as a pollution-sensing torpedo. Still other observations can be done only one at a time, to prevent cables from getting tangled. The resulting Byzantine schedule has, at times, left some researchers on the *Hespérides* working around the clock, sleeping only in short stretches, and others keeping a more comfortable 8 a.m.–3 p.m. schedule.

#### PRACTICE MAKES PERFECT

The success of an expedition can hinge on compatibility — or lack thereof — between team members, so it is important to choose companions carefully, if possible. Wright says that on his field trips in Afar, the most laid-back people cope best with the conditions, and he takes this into account when selecting graduate students and hiring new members of the team. He counsels postdocs and investigators to discuss such considerations before deciding which graduate students or colleagues to invite on their own expeditions.

Researchers who aren’t at liberty to select their field companions should consider arranging a shorter, simpler ‘test’ trip ahead of time, so people can learn to work with each other before the big commitment. Gloria Leon, a psychologist at the University of Minnesota in Minneapolis, says that space agencies often test teams of astronauts for compatibility by taking them on short field excursions or even road trips, exposing them to the tight spaces, tough field conditions and even body odour of a real adventure. Escribano endorses this kind of advance testing. “If I had to screen people for this, I’d lock them in a room and shake it around for a while,” he says. He recommends watching for stress-induced irritability or self-imposed isolation, and offering team members extra attention to help them deal with the stress, if they need it (see “Tips for far-off fieldwork”).

Even if a research group doesn’t have the

## BEFORE YOU GO

### Tips for far-off fieldwork

A few precautions will help to promote an efficient, productive and comfortable research expedition.

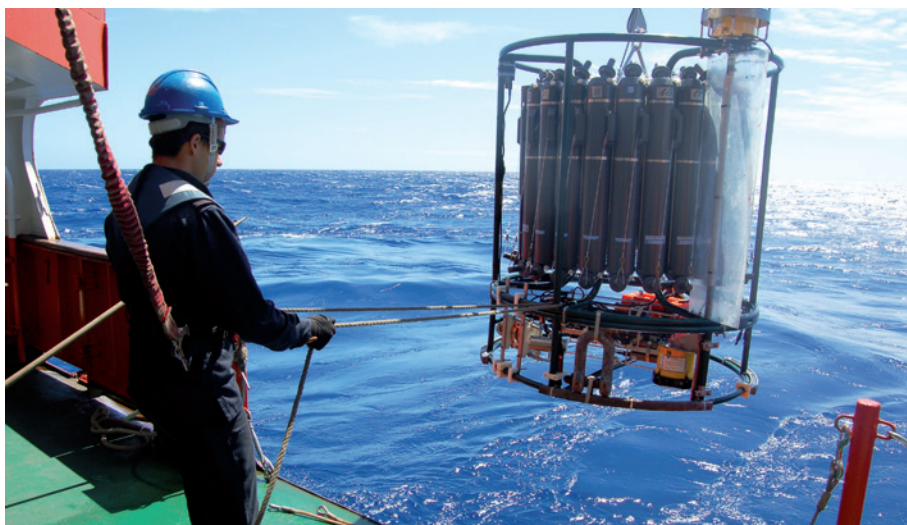
- Cultivate a higher sense of purpose — you’re learning how plankton trap carbon, not just filling dozens of bottles with sea water.
- Make your quarters habitable. People on different shifts shouldn’t have to share the same bunk.
- Test your equipment. The field is not a good place to learn how it works.
- Test your team. Get to know how they will respond to field conditions, and head off any problems that arise.
- Remember to rest your body. In the long run, your work will be better.
- Build downtime into the schedule; if delays don’t use it up, new opportunities will. For example, after the meltdown at the Fukushima Daiichi nuclear plant in Japan in March, the scientists on the Spanish research vessel *BIO Hespérides* began to take previously unplanned measurements of radiation levels.
- Promote open communication to let team members brainstorm or just blow off steam. **L.L.**

luxury of running a trial trip, members can still benefit from working together in advance and trying to sort out any interpersonal difficulties, says Leon. The sole female member of one group of polar explorers that Leon studied complained that the men in the team competed with each other and all vented to her, giving her a disproportionate and unexpected emotional burden. Had she been able to anticipate that dynamic, the explorer might have had a chance to address her companions’ behaviour beforehand and avoid undue strain during the trip.

#### KEEP YOUR PATIENCE

Once a researcher is out in the field, he or she should recognize that not everything is controllable. For example, on one occasion, the *Hespérides* experienced four straight days of rough seas south of Madagascar, and Dachs’s team couldn’t lower the sampling rosette at all. Such things happen on all expeditions. “Sometimes you just can’t collect data,” says Wright. “There’s no point in getting stressed. That isn’t going to make anything happen any faster.”

Focusing on a long-term career or scientific goal can help researchers to cope with the stress of fieldwork, says Leon. She has learnt that the people who succeed on expeditions to the poles — whether they are scientists, sports enthusiasts or explorers — are able to tolerate



Scheduling of data collection involving the sampling rosette caused conflict on the *BIO Hespérides*.

L. LAURSEN





Crew members have to cooperate to overcome difficulties on long voyages.

the conditions because they are trying to achieve something important to them, such as being the first people to ski to the North Pole. Similarly, scientists should keep in mind the scientific problems they will be able to solve, or the papers they will publish, when they're staying up all night to monitor instruments or bathing from a bucket in the desert.

Shrugging off the frustration is easier for old hands than for expedition novices such as Patricia Puerta, a marine-biology graduate student at the CSIC's Mediterranean Institute for Advanced Studies in Majorca, Spain, who is staking her PhD on plankton data collected on the *Hespérides*. On Puerta's first cruise, loose equipment on deck broke her incubation tanks during storms. "My stuff was cannon fodder," she recalls, and she worried that she wasn't making enough progress on her PhD. By the second leg, Puerta had found better ways of rigging her incubation tanks to save them from damage. She also learned the value of patience, she says. "Delays still bother me, but now I'm more accepting." Puerta realized that not every sample was crucial.

Part of becoming an accomplished field-worker is learning to delegate responsibilities such as planning the logistics of an expedition, and to rely on experts to handle things that the researchers can't do as effectively themselves, says Wright. In the Afar, that has meant trusting local scientists or technical staff to help install instruments, negotiate maintenance with Afaris and rent camels for transport. "We couldn't do any of this without our local colleagues," says Wright. "You can't have your eye on all of the balls that are in the air." For Puerta, that meant accepting practical assistance: when her fraction collector broke, preventing her from taking automatic water samples overnight, she didn't, in this instance, have the technical skills to fix it, and had to trust that someone else on board did. "It can make people feel helpless," says Escibano, who, it turned out, travels with

a silver briefcase of electronics equipment and was able to diagnose and repair Puerta's instrument.

Seasoned field scientists must also learn to deal with extended stretches of discomfort and boredom. Breaking the routine on the ship can help to distract people from the discomforts, says Escibano. He has used costume contests and card games to entertain the research team. Others have different ways of dealing with the relentless presence of others, says Leon; she recalls a polar explorer who would "dig real deep in her sleeping bag at night to cover her head" to secure some respite from her companions.

Technology can also mitigate the monotony of fieldwork. Satellites allow team members to ease homesickness by contacting family and friends by phone and through e-mail and social-networking sites, even if only for limited amounts of time. Communications technology can also facilitate scientific decision-making, says Dachs. "On a cruise, you have to react fast, and we use e-mail to consult colleagues on land when we have a problem or a question," he says. From the Indian Ocean, he was able to talk to the other senior expedition planners at research institutes throughout Spain and work out which observations to trim. Not everybody was happy, but everybody was in the loop, and in the end they didn't have to cut quite as many observations as Dachs had feared.

Dachs has one more trick up his sleeve to boost morale: creative time-keeping. The day before arriving in Perth, Australia, for a short break in the cruise, the *Hespérides* approached its offshore meeting point — where a local harbour pilot helps guide the ship into port — earlier than announced. "When you make the cruise schedule," says Dachs, "you always underestimate the ship's speed." ■

**Lucas Laursen** is a freelance journalist based in Zurich, Switzerland.

## UNITED KINGDOM

### Postdocs with promise

The University of Birmingham, UK, has launched a global search for 50 postdocs, mainly in science, engineering and maths. The university is offering five-year fellowships leading to permanent academic posts, providing the stability sought by many young researchers. It is initially investing between £3.5 million (US\$5.7 million) and £5 million in the scheme. Applicants must have achieved a PhD in the past seven years; postdocs will earn about £37,000 annually. "Our vision is to reinvigorate research while creating a cohort of outstanding people that identify with this university," says Adam Tickell, pro-vice-chancellor for research. Tickell adds that if the programme receives many qualified applicants, funding will be available to hire more than the initial 50. The first round of applications closes on 1 September.

## GERMANY

### Research centres open

Twenty-one collaborative research centres in Germany will create up to 630 jobs for postdocs and PhD candidates. The German Research Foundation (DFG), the country's main granting agency, is spending €197 million (US\$289 million) on the centres, which will focus mainly on biological and physical sciences. Research programmes will range from sensory processing to star formation. Opening on 1 July at various host universities, the institutions bring the total number of DFG-funded research centres to 250. The agency will fund the centres for up to 12 years, with about 25 contract scientists each, mostly PhD candidates and the rest postdocs. There are no restrictions on the nationalities of applicants.

## US POSTDOCS

### Women feel isolated

Focus groups convened by the US National Postdoctoral Association (NPA) in Washington DC say that women's academic careers in the United States are hindered by factors including isolation, lack of confidence, perceived lack of status and unfriendly family policies. The groups are part of a National Science Foundation-funded project to identify best practices for advancing female postdocs' careers. Cathleen Johnson Phillips, NPA executive director, says that many women in all disciplines reported inequality and isolation. The NPA will publish its findings in 2012.



Crew members have to cooperate to overcome difficulties on long voyages.

the conditions because they are trying to achieve something important to them, such as being the first people to ski to the North Pole. Similarly, scientists should keep in mind the scientific problems they will be able to solve, or the papers they will publish, when they're staying up all night to monitor instruments or bathing from a bucket in the desert.

Shrugging off the frustration is easier for old hands than for expedition novices such as Patricia Puerta, a marine-biology graduate student at the CSIC's Mediterranean Institute for Advanced Studies in Majorca, Spain, who is staking her PhD on plankton data collected on the *Hespérides*. On Puerta's first cruise, loose equipment on deck broke her incubation tanks during storms. "My stuff was cannon fodder," she recalls, and she worried that she wasn't making enough progress on her PhD. By the second leg, Puerta had found better ways of rigging her incubation tanks to save them from damage. She also learned the value of patience, she says. "Delays still bother me, but now I'm more accepting." Puerta realized that not every sample was crucial.

Part of becoming an accomplished field-worker is learning to delegate responsibilities such as planning the logistics of an expedition, and to rely on experts to handle things that the researchers can't do as effectively themselves, says Wright. In the Afar, that has meant trusting local scientists or technical staff to help install instruments, negotiate maintenance with Afaris and rent camels for transport. "We couldn't do any of this without our local colleagues," says Wright. "You can't have your eye on all of the balls that are in the air." For Puerta, that meant accepting practical assistance: when her fraction collector broke, preventing her from taking automatic water samples overnight, she didn't, in this instance, have the technical skills to fix it, and had to trust that someone else on board did. "It can make people feel helpless," says Escibano, who, it turned out, travels with

a silver briefcase of electronics equipment and was able to diagnose and repair Puerta's instrument.

Seasoned field scientists must also learn to deal with extended stretches of discomfort and boredom. Breaking the routine on the ship can help to distract people from the discomforts, says Escibano. He has used costume contests and card games to entertain the research team. Others have different ways of dealing with the relentless presence of others, says Leon; she recalls a polar explorer who would "dig real deep in her sleeping bag at night to cover her head" to secure some respite from her companions.

Technology can also mitigate the monotony of fieldwork. Satellites allow team members to ease homesickness by contacting family and friends by phone and through e-mail and social-networking sites, even if only for limited amounts of time. Communications technology can also facilitate scientific decision-making, says Dachs. "On a cruise, you have to react fast, and we use e-mail to consult colleagues on land when we have a problem or a question," he says. From the Indian Ocean, he was able to talk to the other senior expedition planners at research institutes throughout Spain and work out which observations to trim. Not everybody was happy, but everybody was in the loop, and in the end they didn't have to cut quite as many observations as Dachs had feared.

Dachs has one more trick up his sleeve to boost morale: creative time-keeping. The day before arriving in Perth, Australia, for a short break in the cruise, the *Hespérides* approached its offshore meeting point — where a local harbour pilot helps guide the ship into port — earlier than announced. "When you make the cruise schedule," says Dachs, "you always underestimate the ship's speed." ■

**Lucas Laursen** is a freelance journalist based in Zurich, Switzerland.

## UNITED KINGDOM

### Postdocs with promise

The University of Birmingham, UK, has launched a global search for 50 postdocs, mainly in science, engineering and maths. The university is offering five-year fellowships leading to permanent academic posts, providing the stability sought by many young researchers. It is initially investing between £3.5 million (US\$5.7 million) and £5 million in the scheme. Applicants must have achieved a PhD in the past seven years; postdocs will earn about £37,000 annually. "Our vision is to reinvigorate research while creating a cohort of outstanding people that identify with this university," says Adam Tickell, pro-vice-chancellor for research. Tickell adds that if the programme receives many qualified applicants, funding will be available to hire more than the initial 50. The first round of applications closes on 1 September.

## GERMANY

### Research centres open

Twenty-one collaborative research centres in Germany will create up to 630 jobs for postdocs and PhD candidates. The German Research Foundation (DFG), the country's main granting agency, is spending €197 million (US\$289 million) on the centres, which will focus mainly on biological and physical sciences. Research programmes will range from sensory processing to star formation. Opening on 1 July at various host universities, the institutions bring the total number of DFG-funded research centres to 250. The agency will fund the centres for up to 12 years, with about 25 contract scientists each, mostly PhD candidates and the rest postdocs. There are no restrictions on the nationalities of applicants.

## US POSTDOCS

### Women feel isolated

Focus groups convened by the US National Postdoctoral Association (NPA) in Washington DC say that women's academic careers in the United States are hindered by factors including isolation, lack of confidence, perceived lack of status and unfriendly family policies. The groups are part of a National Science Foundation-funded project to identify best practices for advancing female postdocs' careers. Cathleen Johnson Phillips, NPA executive director, says that many women in all disciplines reported inequality and isolation. The NPA will publish its findings in 2012.

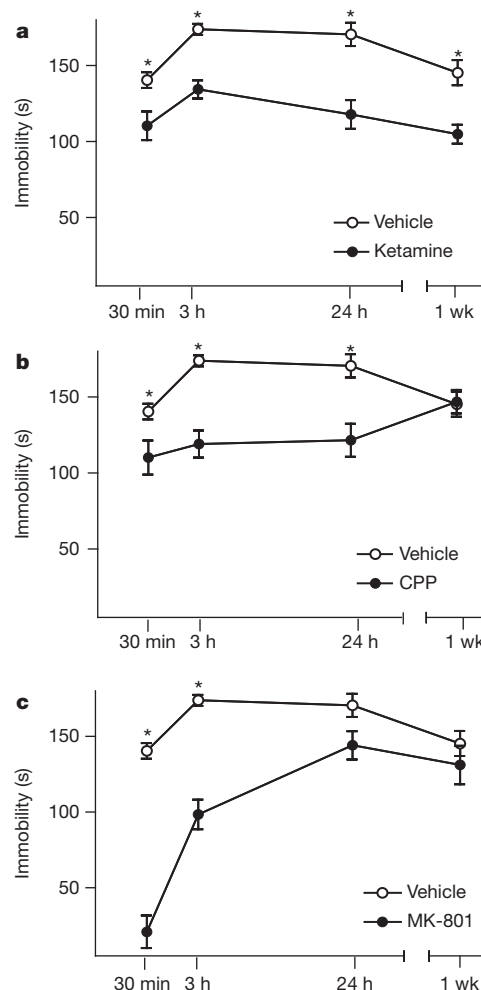


# NMDA receptor blockade at rest triggers rapid behavioural antidepressant responses

Anita E. Autry<sup>1</sup>, Megumi Adachi<sup>1</sup>, Elena Nosyreva<sup>2</sup>, Elisa S. Na<sup>1</sup>, Maarten F. Los<sup>1</sup>, Peng-fei Cheng<sup>1</sup>, Ege T. Kavalali<sup>2</sup> & Lisa M. Monteggia<sup>1</sup>

Clinical studies consistently demonstrate that a single sub-psychomimetic dose of ketamine, an ionotropic glutamatergic NMDAR (*N*-methyl-*D*-aspartate receptor) antagonist, produces fast-acting antidepressant responses in patients suffering from major depressive disorder, although the underlying mechanism is unclear<sup>1–3</sup>. Depressed patients report the alleviation of major depressive disorder symptoms within two hours of a single, low-dose intravenous infusion of ketamine, with effects lasting up to two weeks<sup>1–3</sup>, unlike traditional antidepressants (serotonin re-uptake inhibitors), which take weeks to reach efficacy. This delay is a major drawback to current therapies for major depressive disorder and faster-acting antidepressants are needed, particularly for suicide-risk patients<sup>3</sup>. The ability of ketamine to produce rapidly acting, long-lasting antidepressant responses in depressed patients provides a unique opportunity to investigate underlying cellular mechanisms. Here we show that ketamine and other NMDAR antagonists produce fast-acting behavioural antidepressant-like effects in mouse models, and that these effects depend on the rapid synthesis of brain-derived neurotrophic factor. We find that the ketamine-mediated blockade of NMDAR at rest deactivates eukaryotic elongation factor 2 (eEF2) kinase (also called CaMKIII), resulting in reduced eEF2 phosphorylation and de-suppression of translation of brain-derived neurotrophic factor. Furthermore, we find that inhibitors of eEF2 kinase induce fast-acting behavioural antidepressant-like effects. Our findings indicate that the regulation of protein synthesis by spontaneous neurotransmission may serve as a viable therapeutic target for the development of fast-acting antidepressants.

We examined the acute effect of ketamine in wild-type C57BL/6 mice and detected notable behavioural responses in antidepressant-predictive tasks, including the forced swim test (FST), novelty-suppressed feeding (NSF) and learned helplessness (Supplementary Figs 1a–e and 2a–c)<sup>4</sup>. Ketamine also produced such responses in a sucrose consumption test, as well as in NSF and FST, after chronic mild stress, an animal model of depression (Supplementary Fig. 1f–i). To elucidate the mechanisms underlying the fast-acting antidepressant action of ketamine, we focused on FST, a test that is predictive of non-monoaminergic antidepressant efficacy<sup>4</sup>. We examined the time course of behavioural antidepressant effects in wild-type mice after a single, low-dose treatment with ketamine, (5*S*,10*R*)-(+)-5-methyl-10,11-dihydro-5*H*-dibenzo(a,d)cyclohepten-5,10-imine maleate (MK-801) or 3-((*R*)-2-carboxypiperazin-4-yl)-prop-2-enyl-1-phosphonic acid (CPP) (Fig. 1a–c). After either 30 min or 3 h, each NMDAR antagonist markedly reduced the immobility of mice in FST, when compared to vehicle-treated animals, indicating that NMDAR blockade produces fast-acting antidepressant responses. Notably, in our system, acute treatment with conventional antidepressants did not produce antidepressant-like FST responses (Supplementary Fig. 3), which may require multiple doses<sup>5</sup>. The effects of ketamine and CPP, but not of MK-801, persisted for 24 h (ref. 4) and ketamine's behavioural effect lasted for 1 week. Acute NMDAR-antagonist



**Figure 1 | Time course of NMDAR antagonist-mediated antidepressant-like behavioural effects.** Mean immobility  $\pm$  s.e.m. of C57BL/6 mice in FST after acute treatment with ketamine, CPP or MK-801. Independent groups of mice were used at each time point and for each drug treatment, to avoid behavioural habituation. Analysis of variance (ANOVA)  $F_{(3,27)} = 30.31$ ,  $P < 0.0001$  for treatment groups;  $F_{(3,27)} = 19.06$ ,  $P < 0.0001$  for duration of response;  $F_{(9,81)} = 9.32$ ,  $P < 0.0001$  for treatment-duration interaction. Therefore, we examined treatment effects by time point. **a**, Ketamine ( $3.0 \text{ mg kg}^{-1}$ ) significantly reduced immobility, indicating an antidepressant-like response, at 30 min, 3 h, 24 h and 1 week, compared to vehicle treatment. **b**, CPP ( $0.5 \text{ mg kg}^{-1}$ ) significantly reduced immobility at 30 min, 3 h and 24 h, compared to vehicle treatment. **c**, MK-801 ( $0.1 \text{ mg kg}^{-1}$ ) produced significant decreases in immobility at 30 min and 3 h compared to vehicle treatment.  $n = 10$  mice per group per time point; \*,  $P < 0.05$ . Here and in all figures, error bars represent s.e.m.

<sup>1</sup>Department of Psychiatry, University of Texas Southwestern Medical Center, 5323 Harry Hines Boulevard, Dallas, Texas 75390-9111, USA. <sup>2</sup>Department of Neuroscience, University of Texas Southwestern Medical Center, 5323 Harry Hines Boulevard, Dallas, Texas 75390-9111, USA.

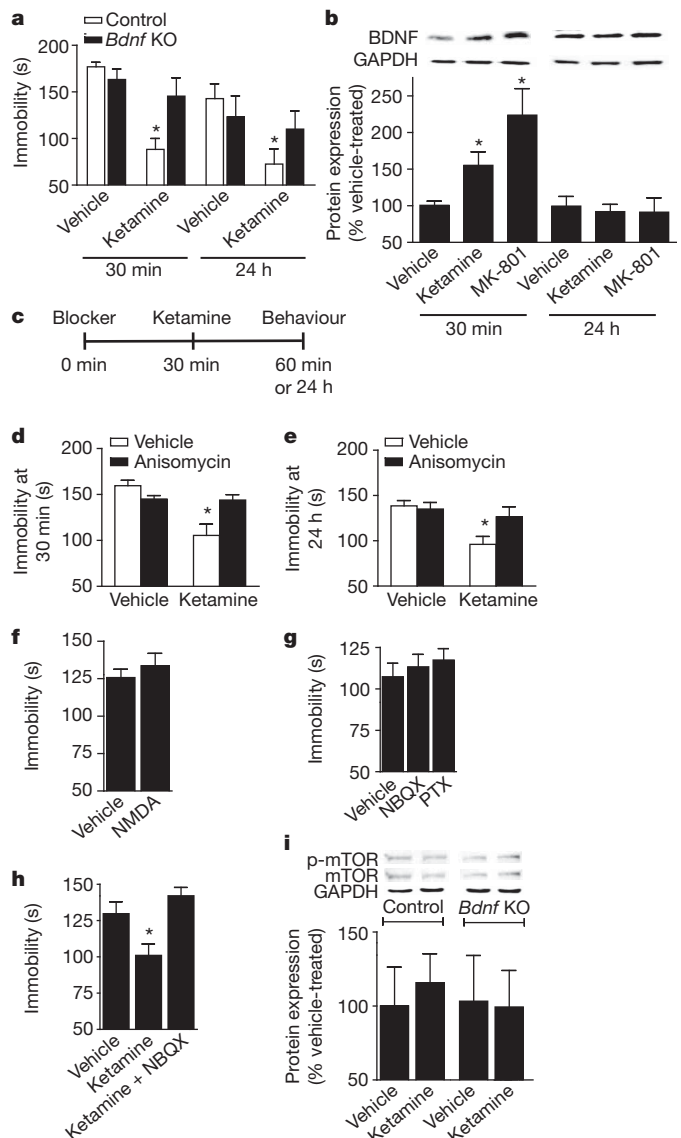


treatment produced no alterations in hippocampal-dependent learning (Supplementary Fig. 1d) or locomotor activity (Supplementary Fig. 4). These drugs have short half-lives (about 2–3 h)<sup>6–8</sup>, indicating that sustained NMDAR-antagonist-induced antidepressant responses are due to synaptic plasticity, not to persistent blockade of receptors.

Brain-derived neurotrophic factor (BDNF) is linked to traditional antidepressant action; BDNF expression in the hippocampus is increased by antidepressants<sup>9</sup> and BDNF deletion in the hippocampus attenuates antidepressant behavioural responses<sup>10–12</sup>. Moreover, intra-ventricular or intrahippocampal BDNF infusion causes rapid, sustained antidepressant-like effects, lasting 3–6 days in FST<sup>13,14</sup>. To examine whether the antidepressant-like response to ketamine is mediated through BDNF, we administered ketamine to inducible *Bdnf*-knockout mice<sup>10</sup>, then observed FST behaviour. After 30 min, ketamine-treated wild-type littermate controls showed significant reductions in immobility, indicating antidepressant-like responses, when compared to vehicle-treated controls (Fig. 2a). However, ketamine did not produce antidepressant-like effects in *Bdnf*-knockouts, indicating that fast-acting antidepressant responses require BDNF. After 24 h, ketamine significantly reduced immobility in controls, but not in *Bdnf* knockouts (Fig. 2a), indicating that ketamine's sustained effects depend on BDNF. To validate this link between NMDAR antagonists and BDNF-mediated antidepressant responses, MK-801 was administered to *Bdnf* knockouts or controls. After 30 min, MK-801 significantly reduced FST immobility in controls, but had no effect in *Bdnf* knockouts (Supplementary Fig. 6). MK-801 did not affect FST behaviour after 24 h (Supplementary Fig. 6), as previously demonstrated (Fig. 1c). We next generated postnatal conditional<sup>15</sup> knockouts in neurotrophic tyrosine kinase receptor, type 2 (*Ntrk2*, also called *TrkB*) and found that these mice were insensitive to ketamine's antidepressant-like effects in FST and NSF (Supplementary Fig. 5a, b). To confirm TrkB engagement, we examined receptor autophosphorylation and found increased TrkB activation after NMDAR antagonist treatment (Supplementary Fig. 5c).

To determine whether NMDAR antagonists alter *Bdnf* expression in the hippocampus, wild-type mice were treated acutely with vehicle, ketamine or MK-801. Quantitative RT-PCR analysis of the coding exon of *Bdnf* showed that *Bdnf* mRNA expression was unaltered by ketamine or MK-801 at either 30 min or 24 h after treatment (Supplementary Fig. 7a). Contrastingly, western blot and ELISA analyses showed that BDNF protein levels were markedly increased at 30 min, but not at 24 h, after NMDAR antagonist treatment (Fig. 2b and Supplementary Fig. 7b). Moreover, the acute effects of ketamine on BDNF extended to its precursor, proBDNF (Supplementary Fig. 7c). These data indicate that rapid increases in BDNF protein translation, not transcription, are necessary for fast-onset antidepressant responses. However, continued BDNF protein upregulation does not underlie ketamine's long-term behavioural effects.

To study further the roles of translation and transcription in ketamine's antidepressant-like effects, we examined FST behaviour in mice treated with the protein synthesis inhibitor anisomycin<sup>16</sup> or with the RNA polymerase inhibitor actinomycin D<sup>17</sup>, which block their respective processes by about 80% within 2 h. We pretreated mice with anisomycin or actinomycin D before treating them with ketamine (Fig. 2c). Anisomycin prevented the ketamine-induced rapid behavioural responses seen at 30 min in FST and NSF paradigms, indicating a dependence on new protein synthesis (Fig. 2d and Supplementary Fig. 8a, b). Anisomycin also prevented ketamine's long-term effect on FST (24 h), indicating that rapid protein translation was involved in sustained antidepressant-like responses (Fig. 2e). We found that the synthesis of both mature BDNF and proBDNF in the hippocampus was sensitive to anisomycin treatment (Supplementary Fig. 8c, d). However, actinomycin D did not affect ketamine's antidepressant-like effect on FST at either time point, indicating that it is independent of new gene expression (Supplementary Fig. 9b, c). To confirm that actinomycin D crossed the blood–brain barrier, we examined *Bdnf* mRNA expression



**Figure 2 | BDNF translation in the antidepressant effects of NMDAR antagonists.** **a**, Immobility in FST after acute treatment with ketamine ( $3.0 \text{ mg kg}^{-1}$ ). At 30 min, ANOVA  $F_{(1,35)} = 17.13$ ,  $P = 0.0002$  for drug;  $F_{(1,35)} = 7.57$ ,  $P = 0.0093$  for genotype–drug interaction; multiple comparisons with *t*-test,  $*$ ,  $P < 0.05$ . At 24 h, in a separate cohort, ANOVA  $F_{(1,29)} = 3.77$ ,  $P = 0.0619$  for treatment; multiple comparisons with *t*-test,  $*$ ,  $P < 0.05$ .  $n = 7$ –12 mice per group. **b**, Densitometric analysis of BDNF (normalized to GAPDH) in the hippocampus after treatment with vehicle (control), ketamine ( $3.0 \text{ mg kg}^{-1}$ ) or MK-801 ( $0.1 \text{ mg kg}^{-1}$ ). At 30 min, ANOVA  $F_{(2,12)} = 6.77$ ,  $P = 0.0108$  for treatment, Bonferroni post hoc test,  $*$ ,  $P < 0.05$ . At 24 h, no significant differences were seen ( $n = 5$ –6 per group). **c**, Protocol for experiments using the blockers anisomycin and actinomycin D. **d**, Immobility at 30 min after anisomycin treatment. ANOVA  $F_{(1,34)} = 11.83$ ,  $P = 0.0016$  for treatment and  $F_{(1,34)} = 10.91$ ,  $P = 0.0023$  for treatment–inhibitor interaction; multiple comparisons,  $*$ ,  $P < 0.05$  ( $n = 8$ –10 per group). **e**, Immobility at 24 h after anisomycin treatment. ANOVA  $F_{(1,31)} = 9.34$ ,  $P = 0.0046$  for treatment; multiple comparisons,  $*$ ,  $P < 0.05$  ( $n = 8$ –10 per group). **f**, Immobility of wild-type mice given vehicle or NMDA ( $75 \text{ mg kg}^{-1}$ ), tested 30 min later in FST. **g**, Immobility of mice given NBQX ( $10 \text{ mg kg}^{-1}$ ) or picrotoxin ( $1 \text{ mg kg}^{-1}$ ), tested 30 min later in FST. **h**, Immobility of mice given vehicle, ketamine ( $3.0 \text{ mg kg}^{-1}$ ) or ketamine + NBQX ( $10 \text{ mg kg}^{-1}$ ) and tested 30 min later in FST. ANOVA  $F_{(2,26)} = 8.226$ ,  $P < 0.0019$ ; Bonferroni post hoc analysis shows that the ketamine effect is reversed by NBQX,  $*$ ,  $P < 0.05$ . **i**, Densitometric analysis of phosphorylated mTOR (normalized to mTOR) in the hippocampus 30 min after treatment with vehicle or ketamine.

in drug-treated animals and found decreased *Bdnf* transcription in the hippocampus (Supplementary Fig. 9a). Taken together, these findings indicate that rapid, transient translation of BDNF is required for ketamine's fast-acting and long-lasting antidepressant-like behavioural effects and that long-term antidepressant responses may be due to alterations in synaptic plasticity, initiated by transient increases in BDNF translation.

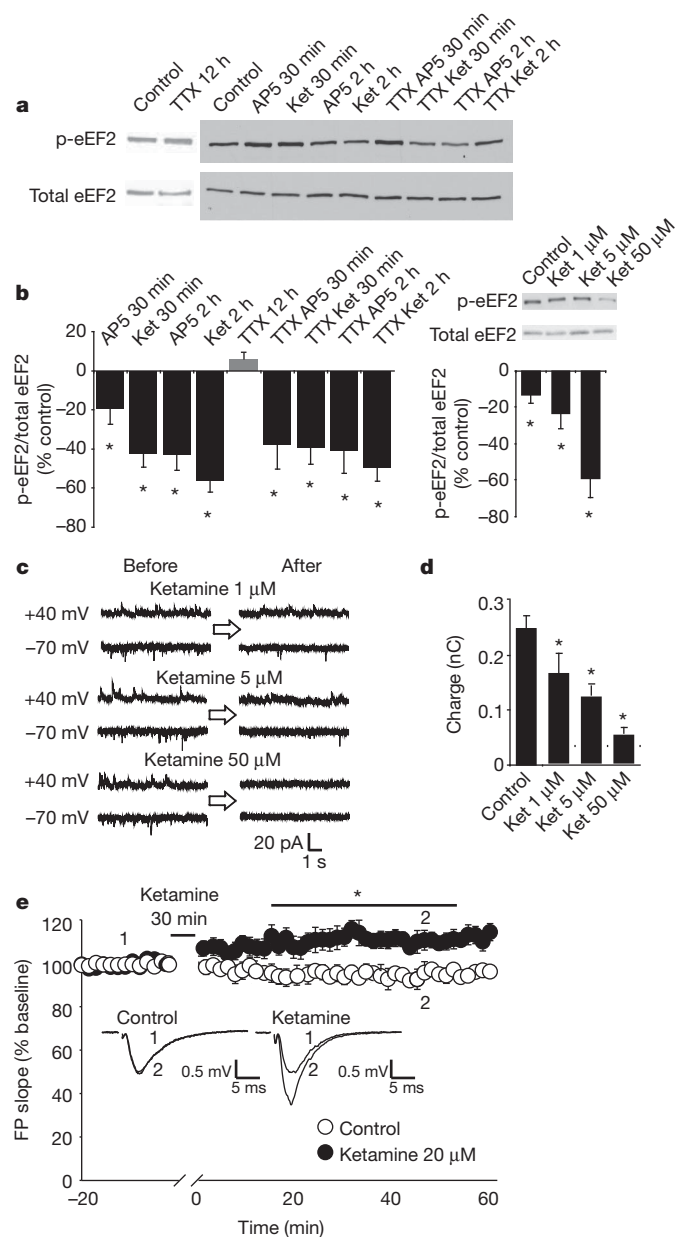
We observed increased levels of BDNF protein in the cortex, but not in the nucleus accumbens, 30 min after acute administration of ketamine or MK-801 (Supplementary Fig. 10a, b). We further investigated whether NMDAR antagonism affected proteins other than BDNF. We found an increased level of activity-regulated cytoskeletal-associated protein (ARC) in the hippocampus (sensitive to anisomycin treatment; Supplementary Fig. 8e) but there was no increase in HOMER or GRIA1 (also known as GLUR1), nor in the phosphorylation of ribosomal protein S6 kinase (Supplementary Fig. 10c–f). Additionally, these proteins remained unaltered in the cortex after acute treatment with NMDAR antagonists (Supplementary Fig. 11a–e).

Synaptic plasticity and ensuing learning processes are often mediated by NMDAR-activation-driven protein translation, but antidepressant-like effects require protein translation induced by NMDAR blockade. To resolve this paradox, we turned to recent evidence that NMDAR blockade by MK-801 or 2-amino-5-phosphonopentanoic acid (AP5) without neuronal activity, augments protein synthesis through eEF2 dephosphorylation (activation). eEF2 is a critical catalytic factor for ribosomal translocation during protein synthesis<sup>18</sup>. In this model, resting NMDAR activity causes sustained activation of eEF2 kinase (eEF2K, or CamKIII), which phosphorylates eEF2, effectively halting translation, whereas acute NMDAR blockade at rest (in the absence of action potentials) attenuates eEF2 phosphorylation, allowing the translation of target transcripts.

To evaluate this model, we tested whether excess synaptic glutamate, possibly elicited by NMDAR blockade, was responsible for the behavioural effects of ketamine. Acute NMDA administration did not alter FST behaviour (Fig. 2f), as previously demonstrated<sup>19</sup>, but it increased ARC expression (Supplementary Fig. 10i), indicating that excess glutamate does not elicit rapid behavioural antidepressant effects. To define the role of neuronal activity in antidepressant behavioural effects, we tested whether NBQX, a blocker of  $\alpha$ -amino-3-hydroxy-5-methyl-4-isoxazole propionic acid (AMPA) channels that reduces neuronal activity, or picrotoxin, a blocker of GABA ( $\gamma$ -aminobutyric acid) channels that increases activity, affected FST behaviour<sup>4,20</sup>. Acute systemic treatment with these drugs did not affect FST behaviour (Fig. 2g) or BDNF synthesis, though picrotoxin enhanced ARC expression in the hippocampus (Supplementary Fig. 10g, h). However, when co-applied with ketamine, NBQX abolished behavioural antidepressant-like responses in FST (Fig. 2h), as previously described<sup>4</sup>. These data indicate that behavioural antidepressant effects are not elicited by alterations in evoked neurotransmission, but require ketamine-mediated augmentation of AMPA-receptor activation.

Recent evidence indicates that cortical mTOR signalling underlies ketamine-mediated antidepressant responses<sup>21</sup>. We investigated whether the rapid behavioural antidepressant effects of ketamine required mTOR activation, and whether this signalling was downstream of BDNF. Regulation of phosphorylated mTOR was not detected after acute administration of ketamine in control or *Bdnf*-knockout hippocampal tissue (Fig. 2i), nor in wild-type cortical tissue (Supplementary Fig. 11d). In earlier work, rapamycin prevented ketamine-mediated antidepressant responses; however, the link between rapamycin and antidepressant-like effects is equivocal<sup>22</sup>. We tested whether pre-treatment with rapamycin could block acute ketamine-mediated FST behaviour. Thirty minutes after ketamine administration, wild-type mice showed antidepressant responses unaffected by rapamycin treatment (Supplementary Fig. 11h). Rapamycin reduced the phosphorylation of ribosomal protein S6 kinase in the cortex and hippocampus (Supplementary Fig. 11f, g), indicating that the rapamycin had penetrated

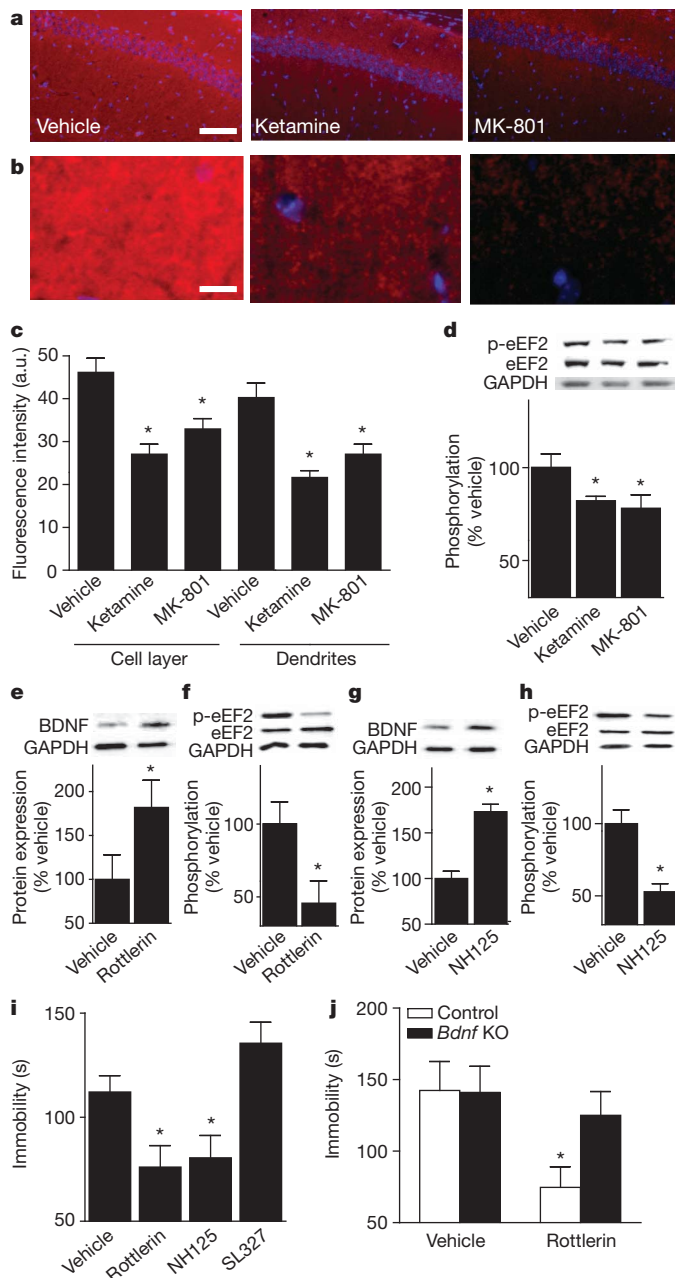
brain tissue. The earlier study examined molecular effects 2 h after drug treatment, or behavioural effects 24 h after drug treatment<sup>21</sup>; therefore mTOR's role in the antidepressant effect of ketamine may be one of maintenance rather than rapid induction.



**Figure 3 | Ketamine blocks NMDAR spontaneous activity, reduces the level of eEF2 phosphorylation and strengthens synaptic responses.**

**a**, Representative Western blots showing eEF2 phosphorylation (p-eEF2) in hippocampal primary cultures. Ket, ketamine; TTX, tetrodotoxin. **b**, Densitometric analysis of p-eEF2 normalized to total eEF2 (left panel). Data are expressed as mean percentage  $\pm$  s.e.m. Tetrodotoxin alone does not alter p-eEF2, whereas AP5 or ketamine, with or without tetrodotoxin, significantly reduce the level of p-eEF2, as assessed by *t*-test (\*,  $P < 0.05$ ). Right panel: application of 1  $\mu$ M, 5  $\mu$ M or 50  $\mu$ M ketamine causes dose-dependent decreases in p-eEF2, as assessed by *t*-test (\*,  $P < 0.05$ ). **c**, Representative traces of NMDAR spontaneous activity after application of 1  $\mu$ M, 5  $\mu$ M or 50  $\mu$ M ketamine. **d**, Quantification of charge transfer (10 s) reveals significant effects, as assessed by *t*-test, for all ketamine concentrations compared to controls ( $n = 6-16$ ; \*,  $P < 0.05$ ). **e**, Field-potential (FP) slopes are plotted as a function of time. Representative field-potential traces (average 2 min) are shown during baseline (1) and at 45 min (2). The asterisk refers to significantly different field-potential values (\*,  $P < 0.05$ ). For statistical analysis, we used two-way repeated ANOVA with Bonferroni post hoc analysis. The drug-time interaction was significant ( $F_{(143,1430)} = 6.723$ ,  $P < 0.001$ ).

To determine whether ketamine inhibits NMDA-receptor-mediated spontaneous miniature excitatory postsynaptic currents (NMDAR-mEPSCs)<sup>3,24</sup> when at rest, and whether it regulates eEF2 phosphorylation, we tested its impact on hippocampal neurons *in vitro*. After ketamine perfusion (1  $\mu$ M, 5  $\mu$ M or 50  $\mu$ M), we recorded NMDAR-mEPSCs (Fig. 3c) and detected a significant decrease within minutes, similar to the effect of AP5 (ref. 23). Moreover, protein extracts from ketamine-treated neurons showed decreased eEF2 phosphorylation compared to vehicle-treated cultures, indicating that ketamine, in the absence of neuronal activity, dose-dependently leads to eEF2 de-phosphorylation, permitting protein synthesis (Fig. 3a, b). Additionally, we evaluated ketamine's effects on hippocampal field potentials. Acute application of ketamine (20  $\mu$ M, at rest) potentiated the synaptic responses subsequently evoked in hippocampal slices (Fig. 3e), further showing that increased AMPA-mediated neurotransmission underlies ketamine's antidepressant-like behavioural effects. This result is consistent with findings regarding BDNF-dependent and protein-synthesis-dependent synaptic plasticity<sup>25</sup>.



To examine whether the fast-acting antidepressant response is mediated via eEF2, we administered ketamine or MK-801 to wild-type mice and analysed eEF2 phosphorylation. Within 30 min, ketamine and MK-801 led to rapid decreases in the level of phosphorylated eEF2 in the hippocampus (Fig. 4a–c and Supplementary Figs 12 and 13), detected by immunostaining and western blot analysis (Fig. 4d). However, cortical levels of phosphorylated eEF2 were unaltered after acute NMDAR-antagonist treatment (Supplementary Fig. 11f). To examine whether eEF2K inhibition alters BDNF protein expression *in vivo*, the eEF2K inhibitors rottlerin or 1-hexadecyl-2-methyl-3-(phenylmethyl)-1H-imidazolium iodide (NH125) were administered to wild-type mice and the mice were killed 30 min later. Rottlerin and NH125 produced significantly increased BDNF protein expression (Fig. 4e, g), with corresponding significant decreases in phosphorylated eEF2 in the hippocampus (Fig. 4f, h). To assess directly whether eEF2K inhibition is sufficient to mediate fast-acting antidepressant-like responses, wild-type mice were treated with rottlerin or NH125 and then examined in FST. Both rottlerin and NH125 produced significant decreases in FST immobility at 30 min (Fig. 4i), a timescale similar to that of the effects of NMDAR antagonists, indicating that fast-acting behavioural effects are mediated through eEF2K inhibition. To test whether mitogen-activated protein kinase (MAPK), a regulator of protein translation during neural activity, affects FST behaviour, we treated wild-type mice with the inhibitor SL327. This treatment reduced MAPK phosphorylation in hippocampal tissue (Supplementary Fig. 10j), but did not affect FST behaviour (Fig. 4i), indicating that antidepressant-like effects are specific to eEF2K inhibition during resting spontaneous glutamatergic signalling. We found that an acute dose of rottlerin or NH125 did not affect locomotor activity, but that antidepressant-related behavioural effects were long-lasting (Supplementary Fig. 14a–f). To validate the finding that antidepressant effects after eEF2K inhibition were mediated through BDNF, we administered rottlerin to *Bdnf*-knockout mice and tested FST behaviour. Like NMDAR antagonists, rottlerin was ineffective in *Bdnf* knockouts, showing that increased *Bdnf* expression upon eEF2K inhibition is required to produce antidepressant-like behavioural responses (Fig. 4j).

Our data support the hypothesis that ketamine produces rapidly acting antidepressant-like behavioural effects through inhibition of spontaneous NMDAR-mEPSCs, leading to decreased eEF2K activity, thus permitting rapid increases in BDNF translation (Supplementary Fig. 15) which may, in turn, exert strong influences on presynaptic or

**Figure 4 | Rapid antidepressant-like behaviour is mediated by decreased p-eEF2 and increased BDNF translation.** **a**, Images of CA1 pyramidal and stratum radiatum layers after acute treatment with vehicle, ketamine or MK-801. Scale bar, 100  $\mu$ m; red, p-eEF2; blue, DAPI. **b**, Magnification of stratum radiatum; scale bar, 20  $\mu$ m. **c**, ImageJ analysis of average fluorescence intensity (a.u., arbitrary units). ANOVA on cell layer,  $F_{(2,23)} = 13.13$ ,  $P = 0.0002$  for treatment; ANOVA on dendrites,  $F_{(2,23)} = 14.06$ ,  $P = 0.0001$  for treatment ( $n = 4$  per group; \*,  $P < 0.05$ ). **d**, Densitometric analysis of p-eEF2 normalized to total eEF2 in the hippocampus after treatment with NMDAR antagonists. ANOVA  $F_{(2,23)} = 3.183$ ,  $P = 0.03$  for treatment ( $n = 8$  per group). **e–h**, Densitometric analyses of BDNF and p-eEF2. Significant increases are seen in hippocampal BDNF protein levels (normalized to GAPDH) with rottlerin (5 mg kg<sup>-1</sup>) versus vehicle (**e**), and with NH125 (5 mg kg<sup>-1</sup>) versus vehicle (**g**) ( $t$ -tests, \*,  $P < 0.05$ ). Significant decreases are seen in p-eEF2 (normalized to total eEF2) with rottlerin versus vehicle (**f**) and NH125 versus vehicle (**h**) ( $t$ -tests, \*,  $P < 0.05$ ). **i**, Immobility in FST of wild-type mice given acute rottlerin (5 mg kg<sup>-1</sup>) or NH125 (5 mg kg<sup>-1</sup>). ANOVA  $F_{(3,44)} = 8.13$ ,  $P = 0.0002$  for treatment; Bonferroni post hoc analysis shows significance with rottlerin or NH125 versus vehicle (\*,  $P < 0.05$ ), but not with the MAPK inhibitor SL327 (10 mg kg<sup>-1</sup>). **j**, Immobility of *Bdnf*-knockout mice or littermate controls given acute rottlerin (5 mg kg<sup>-1</sup>) and tested 30 min later in FST. ANOVA  $F_{(1,19)} = 5.77$ ,  $P = 0.0267$  for treatment; Bonferroni post hoc analysis for rottlerin versus vehicle-treated controls (\*,  $P < 0.05$ ;  $n = 5–7$  per group).



postsynaptic efficacy<sup>26,27</sup>. We found that fast-acting antidepressant-like effects cannot be elicited by disinhibition of behavioural circuitry, or by evoked neurotransmission, but must rely on enhanced neurotransmission after NMDAR-antagonist-induced plasticity, occurring at rest<sup>18</sup>. The observation of behavioural effects mediated through spontaneous neurotransmission provides the first evidence that tonic resting neurotransmission is involved in behaviour, and supports the notion that spontaneous and evoked forms of glutamatergic signalling are segregated<sup>18,23,28,29</sup>. These data demonstrate that eEF2K inhibition, resulting in de-suppression of protein translation, is sufficient to produce antidepressant-like effects, implicating eEF2K inhibitors as potential novel major depressive disorder treatments with rapid onset. Moreover, our results show that synaptic translational machinery may serve as a viable therapeutic target for the development of faster-acting antidepressants.

## METHODS SUMMARY

Behavioural studies were performed using adult male C57BL/6 wild-type or mutant mouse strains, maintained as previously described<sup>10,15</sup>. All drugs were administered via intraperitoneal injection. Antidepressant-like behaviour was assessed using the forced swim test, as previously described<sup>4</sup>. Briefly, animals were placed in a cylinder of water at 22–24 °C for 6 min and immobility was measured during the last 4 min of the test. Molecular studies consisted of western blot analysis or quantitative PCR performed on whole-cell lysates from medial prefrontal cortex or anterior hippocampus. Electrophysiological studies were performed as previously described in cultured neurons (whole-cell recordings<sup>23</sup>) or in hippocampal slices (field recordings<sup>10</sup>).

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

Received 25 October 2010; accepted 21 April 2011.

Published online 15 June 2011.

- Zarate, C. A. Jr *et al.* A randomized trial of an *N*-methyl-D-aspartate antagonist in treatment-resistant major depression. *Arch. Gen. Psychiatry* **63**, 856–864 (2006).
- Berman, R. M. *et al.* Antidepressant effects of ketamine in depressed patients. *Biol. Psychiatry* **47**, 351–354 (2000).
- Price, R. B., Nock, M. K., Charney, D. S. & Mathew, S. J. Effects of intravenous ketamine on explicit and implicit measures of suicidality in treatment-resistant depression. *Biol. Psychiatry* **66**, 522–526 (2009).
- Maeng, S. *et al.* Cellular mechanisms underlying the antidepressant effects of ketamine: role of  $\alpha$ -amino-3-hydroxy-5-methylisoxazole-4-propionic acid receptors. *Biol. Psychiatry* **63**, 349–352 (2008).
- Detke, M. J., Johnson, J. & Lucki, I. Acute and chronic antidepressant drug treatment in the rat forced swimming test model of depression. *Exp. Clin. Psychopharmacol.* **5**, 107–112 (1997).
- Sinner, B. & Graf, B. M. Ketamine. *Handb. Exp. Pharmacol.* **182**, 313–333 (2008).
- Schwartz, P. H. & Wasterlain, C. G. Cardiac arrest and resuscitation alters the pharmacokinetics of MK-801 in the rat. *Res. Commun. Chem. Pathol. Pharmacol.* **73**, 181–195 (1991).
- Kristensen, J. D., Hartvig, P., Karlsten, R., Gordh, T. & Halldin, M. CSF and plasma pharmacokinetics of the NMDA receptor antagonist CPP after intrathecal, extradural and i.v. administration in anaesthetized pigs. *Br. J. Anaesth.* **74**, 193–200 (1995).
- Chen, B., Dowlathshahi, D., MacQueen, G. M., Wang, J. F. & Young, L. T. Increased hippocampal BDNF immunoreactivity in subjects treated with antidepressant medication. *Biol. Psychiatry* **50**, 260–265 (2001).
- Monteggia, L. M. *et al.* Essential role of brain-derived neurotrophic factor in adult hippocampal function. *Proc. Natl Acad. Sci. USA* **101**, 10827–10832 (2004).
- Berton, O. *et al.* Essential role of BDNF in the mesolimbic dopamine pathway in social defeat stress. *Science* **311**, 864–868 (2006).
- Adachi, M., Barrot, M., Autry, A. E., Theobald, D. & Monteggia, L. M. Selective loss of brain-derived neurotrophic factor in the dentate gyrus attenuates antidepressant efficacy. *Biol. Psychiatry* **63**, 642–649 (2008).
- Shirayama, Y., Chen, A. C., Nakagawa, S., Russell, D. S. & Duman, R. S. Brain-derived neurotrophic factor produces antidepressant effects in behavioral models of depression. *J. Neurosci.* **22**, 3251–3261 (2002).
- Hoshaw, B. A., Malberg, J. E. & Lucki, I. Central administration of IGF-I and BDNF leads to long-lasting antidepressant-like effects. *Brain Res.* **1037**, 204–208 (2005).
- Akbarian, S. *et al.* Brain-derived neurotrophic factor is essential for opiate-induced plasticity of noradrenergic neurons. *J. Neurosci.* **22**, 4153–4162 (2002).
- Lattal, K. M. & Abel, T. Different requirements for protein synthesis in acquisition and extinction of spatial preferences and context-evoked fear. *J. Neurosci.* **21**, 5773–5780 (2001).
- Capasso, A., Di Giannuario, A., Loizzo, A., Pieretti, S. & Sorrentino, L. Actinomycin D blocks the reducing effect of dexamethasone on amphetamine and cocaine hypermotility in mice. *Gen. Pharmacol.* **27**, 707–712 (1996).
- Sutton, M. A., Taylor, A. M., Ito, H. T., Pham, A. & Schuman, E. M. Postsynaptic decoding of neural activity: eEF2 as a biochemical sensor coupling miniature synaptic transmission to local protein synthesis. *Neuron* **55**, 648–661 (2007).
- Poleszak, E. *et al.* NMDA/glutamate mechanism of antidepressant-like action of magnesium in forced swim test in mice. *Pharmacol. Biochem. Behav.* **88**, 158–164 (2007).
- Fernandez, F. *et al.* Pharmacotherapy for cognitive impairment in a mouse model of Down syndrome. *Nature Neurosci.* **10**, 411–413 (2007).
- Li, N. *et al.* mTOR-dependent synapse formation underlies the rapid antidepressant effects of NMDA antagonists. *Science* **329**, 959–964 (2010).
- Cleary, C. *et al.* Antidepressant-like effects of rapamycin in animal models: Implications for mTOR inhibition as a new target for treatment of affective disorders. *Brain Res. Bull.* **76**, 469–473 (2008).
- Atasoy, D. *et al.* Spontaneous and evoked glutamate release activates two populations of NMDA receptors with limited overlap. *J. Neurosci.* **28**, 10151–10166 (2008).
- Espinosa, F. & Kavalali, E. T. NMDA receptor activation by spontaneous glutamatergic neurotransmission. *J. Neurophysiol.* **101**, 2290–2296 (2009).
- Tanaka, J. *et al.* Protein synthesis and neurotrophin-dependent structural plasticity of single dendritic spines. *Science* **319**, 1683–1687 (2008).
- Jakawich, S. K. *et al.* Local presynaptic activity gates homeostatic changes in presynaptic function driven by dendritic BDNF synthesis. *Neuron* **68**, 1143–1158 (2010).
- Lindskog, M. *et al.* Postsynaptic GluA1 enables acute retrograde enhancement of presynaptic function to coordinate adaptation to synaptic inactivity. *Proc. Natl Acad. Sci. USA* **107**, 21806–21811 (2010).
- Sutton, M. A. & Schuman, E. M. Partitioning the synaptic landscape: distinct microdomains for spontaneous and spike-triggered neurotransmission. *Sci. Signal.* **2**, pe19 (2009).
- Kavalali, E. T. *et al.* Spontaneous neurotransmission: an independent pathway for neuronal signaling? *Physiology (Bethesda)* **26**, 45–53 (2011).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank M. A. Mahgoub for assistance with the animal experiments, S. Birnbaum and A. Pettersen for assistance with the behavioural testing and members of the Monteggia and Kavalali laboratories for discussions and comments on the manuscript. This work was supported by grant MH070727 (L.M.M.), grant MH066198 (E.T.K.) and the Division of Basic Sciences Training Program at UT Southwestern Medical Center, T32 MH 76690-02 (A.E.A.). E.T.K. is an Established Investigator of the American Heart Association.

**Author Contributions** A.E.A. performed the behavioural experiments. A.E.A., M.A. and M.F.L. contributed to the molecular experiments. E.N. performed the electrophysiology experiments, E.S.N. performed the TrkB behavioural experiments and A.E.A. and P.-f.C. performed the statistical analyses. A.E.A. also made the figures and wrote the corresponding sections of the paper. E.T.K. and L.M.M. designed the study, supervised the experiments and wrote the paper.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at [www.nature.com/nature](http://www.nature.com/nature). Correspondence and requests for materials should be addressed to L.M.M. ([lisa.monteggia@utsouthwestern.edu](mailto:lisa.monteggia@utsouthwestern.edu)) or E.T.K. ([ege.kavalali@utsouthwestern.edu](mailto:ege.kavalali@utsouthwestern.edu)).

## METHODS

**Mouse.** C57BL/6 male mice aged 6–8 weeks old were habituated to animal facilities for 1 week before behavioural testing. Mice were kept on a 12 h/12 h light/dark cycle and were given access to food and water *ad libitum*. Inducible *Bdnf* knockouts were generated from a trigenic cross of NSE-tTA, TetOp-Cre and floxed *Bdnf* mice, as previously described<sup>10</sup>. Conditional *Ntrk2*-knockout mice were made by crossing CamK-cre(93) (ref. 15) to floxed *Ntrk2* mice. For all behavioural testing, male mice were 2–4 months old and weight-matched, and groups were balanced by genotype. All animal procedures conformed to the guide for the care and use of laboratory animals and were approved by the institutional animal care and use committee at UT Southwestern Medical Center.

**Drugs.** All drugs were injected intraperitoneally. Concentrations were as follows: ketamine (Fort Dodge Animal Health) 3.0 mg kg<sup>-1</sup>, MK-801 (Sigma) 0.1 mg kg<sup>-1</sup> and CPP (Sigma) 0.5 mg kg<sup>-1</sup> in 0.9% saline; anisomycin (Sigma) 100 mg kg<sup>-1</sup> (dissolved in HCl/saline, final pH 7.4); actinomycin D (Sigma) 0.5 mg kg<sup>-1</sup> in 5% ethanol; roflumetinol and NH125 (Sigma) 5 mg kg<sup>-1</sup> in 20–100% DMSO; SL327 (Sigma) 10 mg kg<sup>-1</sup> in 100% DMSO<sup>30</sup>; NMDA (Sigma) 75 mg kg<sup>-1</sup>, NBQX (Sigma) 10 mg kg<sup>-1</sup> and picrotoxin (Sigma) 1 mg kg<sup>-1</sup> in 0.9% saline; rapamycin (Sigma) 1.0 mg kg<sup>-1</sup> dissolved in 50% DMSO.

**Sucrose consumption test.** Group-housed mice were habituated to a 1% solution of sucrose in tap-water for 48 h. The mice were then habituated to water-deprivation periods of 4 h, 14 h and 19 h, followed by a 1 h exposure to the sucrose solution for 3 days with intervening access to normal drinking water. To assess individual sucrose intake, the group-housed mice were water-deprived overnight and then housed temporarily in a new cage. Each test mouse was placed in its home cage for 1 h with access to the 1% sucrose solution. The bottle of sucrose solution was weighed before and after the test to determine sucrose intake. A water test was performed in a similar manner the following day. Data are expressed as a percentage of sucrose to total volume consumed in both sucrose and water trials.

**Elevated plus maze.** Mice were placed in the centre of a plus maze (each arm 33 cm × 5 cm) that was elevated 1 m above the floor with two open arms and two closed arms (25-cm-tall walls on the closed arms) at 40 lx. The exploratory activity was monitored for 5 min with a video tracking system and the duration, in seconds, spent in the closed and open arms was recorded by EthoVision software.

**Novelty-suppressed feeding.** Briefly, group-housed animals were food-deprived for 24 h and then placed in a temporary home cage for 30 min. For the test, individual mice were placed in a 42 × 42 cm open-field arena at 40 lx. A single pellet of the mouse's normal food chow was placed in the centre of the open-field arena. Each animal was placed in a corner of the arena and allowed to explore for up to 10 min. The trial ended when the mouse chewed a part of the chow. The amount of food consumed in the home cage was taken as the weight of chow consumed in 5 min, as a control measure for appetite.

**Context and cued fear conditioning.** Fear conditioning was performed as previously described<sup>5</sup>. Briefly, mice were placed in individual chambers for 2 min, followed by a loud tone (90 dB) for 30 s, immediately followed by a 0.5 mA foot-shock for 2 s. After 1 min, mice received a second pairing of tone and footshock, as described. Mice were placed in home cages until 24 h later, when the mice were placed back in the same boxes without a tone or shock. The amount of time that the animal spent freezing was scored by an observer blind to genotype. Freezing behaviour was defined as no movement except for respiration. Four hours later, mice were placed in a novel environment with no tone or shock for 3 min, followed by 3 min of the tone to assess cue-dependent fear conditioning. Again, time spent freezing was recorded as described<sup>10</sup>.

**Learned helplessness.** Mice were trained on one side of a two-chamber shuttlebox (MedAssociates) with the door closed for 1 h, receiving 120 variable-interval shocks (18–44 s, average 30 s; 0.35 mA for 2 s) on 2 training days. On the test day, the door was raised at the onset of the shock and the shock ended either when the mouse stepped through to the other side of the shuttlebox or after 25 s. Latency to step through the door and the number of escape failures were recorded for 15 trials.

**Locomotor activity.** Mice were placed in cages and locomotor activity was recorded for 1 h under red light by photocell beams linked to computer acquisition software (San Diego Instruments).

**Forced swim test.** The forced swim test (FST) was performed as previously described<sup>12</sup>. This test is sensitive to conventional antidepressant treatment<sup>31</sup> as well as to non-monoaminergic antidepressants<sup>4</sup>. Mice were placed for 6 min in a 4 l Pyrex glass beaker containing 3 l of water at 24 ± 1 °C. Water was changed between subjects. All test sessions were recorded by a video camera positioned on the side of the beaker. The videotapes were analysed and scored by an observer blind to group assignment during the last 4 min of the 6 min trial. A decrease in immobility time indicates an antidepressant-like response.

**Chronic mild stress.** Stressed mice were subjected to two randomly selected mild stressors per day, of variable duration (1–12 h), for 28 days. Stressors included

water deprivation, 45° cage-tilt, food deprivation, exposure to rat faeces, cage overcrowding, wet bedding, overnight illumination, dark exposure during normal light cycle, cold bedding, acoustic disturbance (120 dB), strobe lights and cage-mate rotation. Stressors were not applied within 8 h of behavioural testing.

**Time course experiments.** Separate cohorts of C57BL/6 adult male mice were injected intraperitoneally with vehicle or the NMDAR antagonists ketamine (3.0 mg kg<sup>-1</sup>), MK-801 (0.1 mg kg<sup>-1</sup>) or CPP (0.5 mg kg<sup>-1</sup>) at 30 min, 3 h, 24 h or 1 week before FST (*n* = 10 per group). The drug doses were chosen on the basis of previous literature demonstrating an antidepressant-like response in mouse models<sup>4</sup>.

**Anisomycin and actinomycin D experiments.** Separate cohorts of C57BL/6 adult male mice were injected intraperitoneally with either vehicle or anisomycin (100 mg kg<sup>-1</sup>), or with either saline or actinomycin D (0.5 mg kg<sup>-1</sup>), 1 h before FST. Thirty minutes before testing, mice received either a saline or a ketamine injection (3.0 mg kg<sup>-1</sup>) (*n* = 10 per group). For 24 h experiments, mice were given anisomycin (100 mg kg<sup>-1</sup>) or saline 30 min before an injection of ketamine and were tested in the FST 1 day later.

**Inducible *Bdnf*-knockout experiments.** Separate cohorts of inducible *Bdnf* knockout adult male mice and wild-type littermate controls were subjected to FST either 30 min or 24 h after injection with saline, ketamine (3.0 mg kg<sup>-1</sup>) or MK-801 (0.1 mg kg<sup>-1</sup>) (*n* = 7–12 per group).

**Quantitative RT-PCR.** Fresh frozen anterior hippocampal slices (2 per mouse, ~1 mm thick) were dissected and total RNA was extracted using Trizol reagent (Invitrogen), according to manufacturer's instructions. Conditions for cDNA synthesis, amplification and primer sequences were described previously<sup>12</sup>. The fold-change in *Bdnf* expression (coding exon) was normalized to GAPDH.

**Protein quantification.** Anterior hippocampal slices (2 per mouse, ~1 mm thick) were dissected from C57BL/6 mice that had received saline vehicle, ketamine (3.0 mg kg<sup>-1</sup>) or MK-801 (0.1 mg kg<sup>-1</sup>) injections, either 30 min or 24 h after injection. The slices were rapidly frozen and lysed in buffer containing protease inhibitors and phosphatase inhibitors. Total protein concentration was quantified by Bradford analysis. BDNF quantification was carried out by SDS-polyacrylamide gel electrophoresis. Primary antibodies for BDNF (Santa Cruz Biotechnology) and GAPDH (Cell Signaling) were used at dilutions of 1:200 and 1:10,000, and anti-rabbit secondary antibodies were used at 1:2,000 and 1:50,000, respectively. To measure phosphorylated eEF2 (p-eEF2, Thr 56) and total eEF2, primary antibodies were used at dilutions of 1:1,000 and anti-rabbit secondary antibodies were used at 1:2,000. Mouse anti-ARC (C7, Cell Signaling) was used at a primary dilution of 1:1,000 and secondary dilution of 1:2,000. Phospho-mTOR and total mTOR (Cell Signaling) were both used at primary dilutions of 1:500 and secondary dilutions of 1:10,000. GluR1 (Chemicon) was used at a primary dilution of 1:5,000 and secondary dilution, 1:2,000. Pan-HOMER antibody (Cell Signaling) was used at 1:5,000 with 1:2,000 dilutions for primary and secondary, respectively. Phospho-s6 kinase and total s6 kinase antibodies were used at 1:200 and 1:5,000 for primary dilutions, respectively, and both had secondary dilutions of 1:5,000 (Cell Signaling). Phospho-MAPK and total MAPK antibodies (Cell Signaling) were used at primary dilutions of 1:10,000 and 1:500 respectively and both had secondary dilutions of 1:2,000. Bands developed with enzymatic chemiluminescence (ECL) were exposed to film and films were analysed by ImageJ. BDNF was normalized to GAPDH bands, and p-eEF2 and total eEF2 bands were taken as a ratio of GAPDH-normalized values.

**Immunohistochemistry.** C57BL/6 mice were treated intraperitoneally with saline, ketamine (3.0 mg kg<sup>-1</sup>) or MK-801 (0.1 mg kg<sup>-1</sup>) and killed 30 min later. The protocol is adapted from a previous study<sup>32</sup>. Brains were fresh-dissected and fixed for 72 h in ice-cold 4% paraformaldehyde. Brains were cryoprotected for 2 or more hours in 20% glycerol, sectioned on a freezing microtome at 30 µm and preserved in 1× PBS with 0.01% sodium azide. Floating sections were washed in 2× SSC, followed by antigen-unmasking in 50:50 acetone:methanol, performed at 4 °C. Sections were rinsed and endogenous peroxidase activity was quenched in 1% H<sub>2</sub>O<sub>2</sub> for 30 min. Sections were rinsed in 2× SSC with 0.05% Tween-20. Tissue was blocked for 30 min in 3% normal goat serum diluted in 2× SSC/0.05% Tween, followed by primary antibody, rabbit anti-p-eEF2 (diluted 1:100 in blocking solution; Cell Signaling Technology), and incubation for 48 h at 4 °C. After rinsing in 2× SSC, a horseradish-peroxidase-labelled secondary antibody at 1:200 was applied and the signal was amplified using the tyramide amplification signal system (Perkin Elmer). Slides were counterstained with 4',6'-diamidino-2-phenylindole (DAPI), mounted on superfrost plus slides, dried for 2 h and mounted in DPX mountant.

**ELISA.** A high-sensitivity enzyme-linked immunosorbent assay was used to assess BDNF levels, as per manufacturer's instructions (Promega). Briefly, hippocampal lysates were prepared in the recommended buffer, diluted 1:4 in 1× PBS and acid-treated as instructed by the manufacturer. A 96-well plate (Nunc) was coated overnight in carbonate coating buffer, blocked in the provided sample buffer for

2 h at 26 °C and treated with recombinant human BDNF antibody for 2 h at 26 °C. Acid-treated samples and provided standards were added to the plate in duplicate. Wells were then treated for 1 h at RT with anti-IgY conjugated to horseradish peroxidase and colour was developed with the provided 3,3',5,5'-tetramethylbenzidine (TMB) solution for 10 min, then stopped with 1 M HCl. Absorbance of wells was measured at 450 nm. BDNF concentration was determined by comparing the mean absorbance of the duplicate samples to the standards. BDNF concentration was then normalized to total protein content and expressed as pg of BDNF per µg of total protein.

**Cell culture.** Dissociated hippocampal cultures were prepared as previously described<sup>33</sup>. Briefly, whole hippocampi were dissected from Sprague–Dawley rats at postnatal day 0–3. Tissue was trypsinized (10 mg ml<sup>-1</sup> trypsin) for 10 min at 37 °C, mechanically dissociated by pipetting and plated on Matrigel-coated coverslips. Cytosine arabinoside (4 µM, Sigma) was added at day 1 *in vitro* and the concentration of cytosine arabinoside was reduced to 2 µM at day 4 *in vitro*. All experiments were performed on cultures at day 14–21.

**Cell culture recordings.** Whole-cell patch-clamp recordings were performed on hippocampal pyramidal neurons. Data were acquired using a MultiClamp 700B amplifier and Clampex 9.0 software (Molecular Devices). Recordings were filtered at 2 kHz and sampled at 200 µs. A modified Tyrode's solution containing 150 mM NaCl, 4 mM KCl, 2 mM MgCl<sub>2</sub>, 2 mM CaCl<sub>2</sub>, 10 mM glucose and 10 mM HEPES, pH 7.4, was used as external bath solution. The pipette-internal solution contained 115 mM CsMeSO<sub>3</sub>, 10 mM CsCl, 5 mM NaCl, 10 mM HEPES, 0.6 mM EGTA, 20 mM tetraethylammonium chloride, 4 mM Mg-ATP, 0.3 mM Na<sub>3</sub>GTP, pH 7.35, and 10 mM QX-314 (*N*-(2,6-dimethylphenylcarbamoylmethyl)-triethylammonium bromide), 300 mosM. Series resistance was 10–30 mΩ. To record and isolate NMDAR-mEPSCs, the MgCl<sub>2</sub> concentration was reduced to 0.1 mM and 2,3-dihydroxy-6-nitro-7-sulfamoyl-benzo(f)quinoxaline-2,3-dione (NBQX; 10 µM, Sigma) and picrotoxin (50 µM; Sigma) were added to the bath solution to block AMPA-receptor-mediated excitatory currents and GABA (γ-aminobutyric acid) receptor-mediated inhibitory currents, respectively. The baseline for the analysis of NMDAR-mEPSCs was automatically determined as the average current level of silent episodes during a recording. The events were selected at a minimum threshold of 4 pA and the area under current deflection was calculated to quantify charge transfer<sup>18</sup>.

**Field recordings.** Field recordings were made from hippocampal slices from Sprague–Dawley rats obtained from Charles River Laboratories. Slices (400 µm) were prepared from rats at 15–25 days old. Rats were anesthetized with euthasol

(50 mg kg<sup>-1</sup>) and decapitated soon after the disappearance of corneal reflexes. The brain was removed, dissected and then sliced using a vibratome (1000 Plus) in ice-cold dissection buffer containing 2.6 mM KCl, 1.25 mM NaH<sub>2</sub>PO<sub>4</sub>, 26 mM NaHCO<sub>3</sub>, 0.5 mM CaCl<sub>2</sub>, 5 mM MgCl<sub>2</sub>, 212 mM sucrose and 10 mM dextrose. Area CA3 was surgically removed from each slice immediately after sectioning. The slices were transferred into a reservoir chamber filled with artificial cerebrospinal fluid (ACSF) containing 124 mM NaCl, 5 mM KCl, 1.25 mM NaH<sub>2</sub>PO<sub>4</sub>, 26 mM NaHCO<sub>3</sub>, 2 mM CaCl<sub>2</sub>, 2 mM MgCl<sub>2</sub> and 10 mM dextrose. Slices were allowed to recover for 2–3 h at 30 °C. ACSF and dissection buffer were equilibrated with 95% O<sub>2</sub> and 5% CO<sub>2</sub>. For recording, slices were transferred to a submerged recording chamber, maintained at 30 °C, and perfused continuously with ACSF at a rate of 2–3 ml min<sup>-1</sup>. Field potentials were recorded with extracellular recording electrodes (1 MΩ) filled with ACSF and placed in the stratum radiatum of area CA1. Field potentials were evoked by monophasic stimulation (duration, 200 µs) of Schaffer collateral/commissural afferents with a concentric bipolar tungsten-stimulating electrode (Frederick Haer Company). Stable baseline responses were collected every 30 s using a stimulation intensity of 10–30 µA, yielding 50–60% of the maximal response. After recording 20 min of stable baseline, the stimulation was stopped and 20 µM ketamine was applied for 30 min, then stimulation was resumed. Field potentials were filtered at 2 kHz, acquired and digitized at 10 kHz on a personal computer using custom software (LabVIEW, National Instruments). Synaptic strength was measured as the initial slope (10–40% of the rising phase) of the field potential. The group data were analysed as follows: (1) the initial slopes of the field potential were expressed as percentages of the preconditioning baseline average; (2) the timescale in each experiment was converted to the time from the end of ketamine application; and (3) the time-matched, normalized data were averaged across experiments.

30. Duman, C. H., Schlesinger, L., Kodama, M., Russell, D. S. & Duman, R. S. A role for MAP kinase signalling in behavioral models of depression and antidepressant treatment. *Biol. Psychiatry* **61**, 661–670 (2007).
31. Porsolt, R. D., Le Pichon, M. & Jalfre, M. Depression: A new animal model sensitive to antidepressant treatments. *Nature* **266**, 730–732 (1977).
32. Park, S. *et al.* Elongation factor 2 and fragile X mental retardation protein control the dynamic translation of Arc/Arg3.1 essential for mGluR-LTD. *Neuron* **59**, 70–83 (2008).
33. Kavalali, E. T., Klingauf, J. & Tsien, R. W. Activity-dependent regulation of synaptic clustering in a hippocampal culture system. *Proc. Natl Acad. Sci. USA* **96**, 12893–12900 (1999).



# Structural basis of steroid hormone perception by the receptor kinase BRI1

Michael Hothorn<sup>1</sup>, Youssef Belkhadir<sup>1,2†</sup>, Marlene Dreux<sup>3</sup>, Tsegaye Dabi<sup>1,2</sup>, Joseph. P. Noel<sup>2,4</sup>, Ian A. Wilson<sup>5,6</sup> & Joanne Chory<sup>1,2</sup>

Polyhydroxylated steroids are regulators of body shape and size in higher organisms. In metazoans, intracellular receptors recognize these molecules. Plants, however, perceive steroids at membranes, using the membrane-integral receptor kinase BRASSINOSTEROID INSENSITIVE 1 (BRI1). Here we report the structure of the *Arabidopsis thaliana* BRI1 ligand-binding domain, determined by X-ray diffraction at 2.5 Å resolution. We find a superhelix of 25 twisted leucine-rich repeats (LRRs), an architecture that is strikingly different from the assembly of LRRs in animal Toll-like receptors. A 70-amino-acid island domain between LRRs 21 and 22 folds back into the interior of the superhelix to create a surface pocket for binding the plant hormone brassinolide. Known loss- and gain-of-function mutations map closely to the hormone-binding site. We propose that steroid binding to BRI1 generates a docking platform for a co-receptor that is required for receptor activation. Our findings provide insight into the activation mechanism of this highly expanded family of plant receptors that have essential roles in hormone, developmental and innate immunity signalling.

Signal perception at the cell surface, and transduction of this signal to the cell's interior, are essential to all life forms. Plants have met this challenge in part by evolving membrane-integral receptor kinases. Many of these receptors are composed of an extracellular leucine-rich repeat (LRR) module and a cytoplasmic kinase domain, connected by a single membrane-spanning helix<sup>1</sup>. Receptors with this architecture (LRR-receptor kinase, LRR-RK), for example, regulate plant growth<sup>2</sup>, development<sup>3,4</sup> and interactions with the environment<sup>5,6</sup>. Their corresponding ligands range from small molecules<sup>7</sup> and peptides<sup>8</sup> to entire proteins<sup>5</sup>.

The LRR-RK BRI1 (refs 2, 9) controls a steroid signalling pathway essential for plant growth<sup>10</sup>. Whereas animal steroid receptors are found predominantly in the nucleus<sup>11</sup>, BRI1 is localized at the plasma-membrane and in endosomes<sup>12</sup>. The following model for BRI1 activation has been proposed: in the absence of brassinosteroid, BRI1's kinase domain is kept in a basal state by its auto-inhibitory carboxy-terminal tail<sup>13</sup>, as well as by interaction with the inhibitor protein BKI1 (ref. 14). Hormone binding to the extracellular domain of BRI1 (refs 7, 15), in a region that includes a ~70 amino acid 'island' domain<sup>16</sup>, causes a change in the receptor (a conformational change in a preformed homodimer<sup>13</sup> or receptor dimerization), leading to autophosphorylation of the BRI1 kinase domain<sup>17</sup>, release of its C-terminal tail<sup>13</sup>, and trans-phosphorylation of the inhibitor BKI1 (refs 14, 18). BKI1 then dissociates from the membrane, allowing BRI1 to interact with a family of smaller LRR-RKs<sup>19</sup>, including the BRI1 ASSOCIATED KINASE 1 (BAK1)<sup>20,21</sup>. The kinase domains of BRI1 and BAK1 trans-phosphorylate each other on multiple sites<sup>22</sup>, and the fully activated receptor triggers downstream signalling events<sup>23</sup>, resulting in major changes in nuclear gene expression<sup>10</sup>.

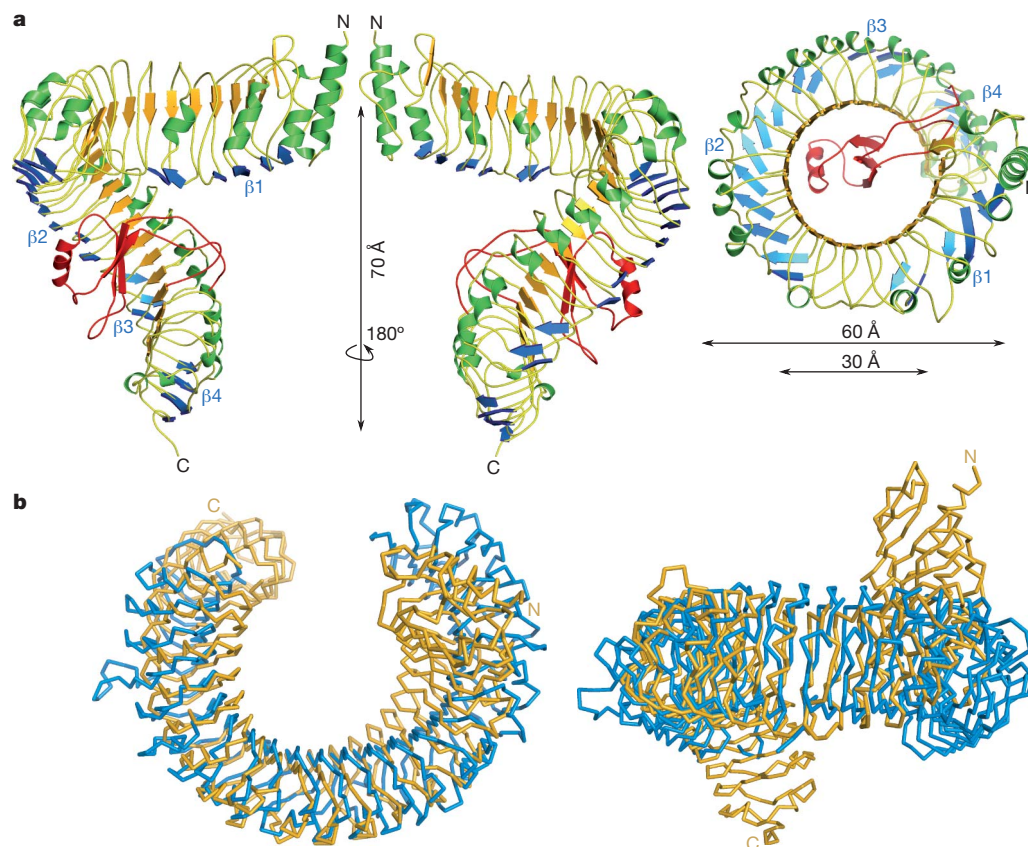
The architecture of BRI1 is reminiscent of animal Toll-like innate immunity receptors (TLRs), and notably several plant LRR-RKs are immunity receptors<sup>5</sup>. It was, thus, reasonable to assume that the BRI1 ectodomain would form a TLR-like horseshoe structure<sup>24</sup> that binds

its ligand along a dimer interface, as observed in several TLRs<sup>25,26</sup>. Here we report the structure of the ligand-binding domain of *Arabidopsis* BRI1 in its free form and bound to the steroid brassinolide, and show that BRI1 folds into a superhelical assembly, whose interior provides the hormone-binding site. Comparison of the free and hormone-bound structures, combined with genetic data, suggests a novel activation mechanism for BRI1 that is distinct from TLRs.

## Overall structure of the BRI1 ectodomain

BRI1 was expressed in baculovirus-infected insect cells and the secreted ectodomain was purified by tandem-affinity and size-exclusion chromatography. The crystal structure was solved to 2.5 Å resolution by single isomorphous replacement (see Methods, Supplementary Table 1 and Fig. 1). BRI1 does not adopt the anticipated TLR-horseshoe structure but forms a right-handed superhelix composed of 25 LRRs (Fig. 1a). The helix completes one full turn, with a rise of ~70 Å. The concave surface, which determines the curvature of the solenoid<sup>27</sup>, is formed by  $\alpha$ - and  $3_{10}$  helices (green in Fig. 1a) that produce inner and outer diameters of ~30 and ~60 Å, respectively. The overall curvature of BRI1 is similar to that of TLR3 (ref. 24; Fig. 1b), but, whereas the TLR3 ectodomain is essentially flat, BRI1 is highly twisted (Fig. 1b). Such twisted assemblies of LRRs have been observed previously with bacterial effector<sup>28</sup> and adhesion proteins<sup>29</sup>, and with the plant defence protein PGIP<sup>30</sup> (Supplementary Fig. 2). The twist of PGIP's LRR domain is caused by a non-canonical, second  $\beta$ -sheet that is oriented perpendicular to the central  $\beta$ -sheet forming the inner surface of the solenoid<sup>30</sup>. Additional  $\beta$ -sheets are also present in our structure (blue in Fig. 1a, Supplementary Fig. 3), but in the case of the much larger BRI1 ectodomain result in a superhelical assembly (Fig. 1a). The second  $\beta$ -strand in PGIP and in BRI1 is followed by an Ile-Pro spine that runs along the outer surface of the helix and provides packing interactions between consecutive LRRs (Fig. 2a and ref. 30). Both structural features are directly linked to

<sup>1</sup>Plant Biology Laboratory, The Salk Institute for Biological Studies, 10010 North Torrey Pines Road, La Jolla, California 92037, USA. <sup>2</sup>Howard Hughes Medical Institute, The Salk Institute for Biological Studies, 10010 North Torrey Pines Road, La Jolla, California 92037, USA. <sup>3</sup>Department of Immunology and Microbial Science, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, California 92037, USA. <sup>4</sup>Jack H. Skirball Center for Chemical Biology and Proteomics, The Salk Institute for Biological Studies, 10010 North Torrey Pines Road, La Jolla, California 92037, USA. <sup>5</sup>Department of Molecular Biology, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, California 92037, USA. <sup>6</sup>Skaggs Institute for Chemical Biology, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, California 92037, USA. <sup>†</sup>Present address: Moroccan Foundation for Advanced Science, Innovation and Research, Biotechnology Development Center, Technopolis Rabatshore, Sala al Jadida, 11000, Morocco.

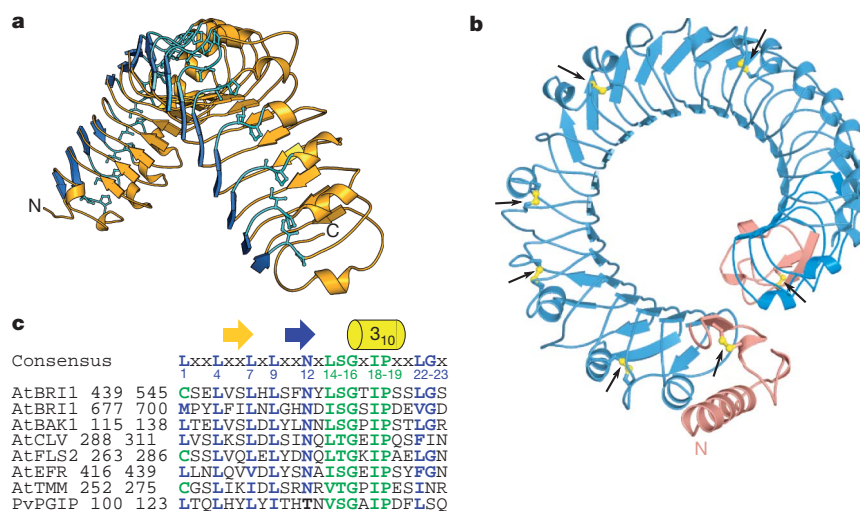


**Figure 1 | The BRI1 ectodomain forms a superhelix.** **a**, Ribbon diagrams of the ectodomain. Front, back and top views are shown (left, middle and right, respectively). The canonical  $\beta$ -sheet is shown in orange, helices in green, the plant-specific  $\beta$ -sheets in blue, and the island domain is depicted in red.

**b**, Structural comparison of BRI1 ( $C_\alpha$  trace, in yellow) and TLR3 (in blue; PDB 1ZIW)<sup>24</sup>. The structures superimpose with an root mean square deviation of 4.2 Å between 341 corresponding  $C_\alpha$  atoms. Top and side views are shown (left and right, respectively). The island domain has been omitted for clarity.

the Lt/sGxIP consensus sequence of the plant-specific LRR subfamily<sup>31</sup> (Fig. 2c, Supplementary Fig. 4 and Supplementary Table 2). Because this consensus sequence is found in other plant receptor kinases, these receptors may also harbour twisted LRR domains (Fig. 2c), making BRI1 the primary template for the study of diverse signalling pathways in plants<sup>3–6</sup>.

Amino- and C-terminal flanking regions that cap the hydrophobic core of the BRI1 solenoid are similar to caps previously described for PGIP<sup>30</sup> (Supplementary Fig. 5). Notably, not only are these caps stabilized by disulphide bridges, but five additional disulphide bonds link consecutive LRR segments in the N-terminal half of the BRI1 ectodomain (Fig. 2b, Supplementary Fig. 4 and Supplementary Table 2).



**Figure 2 | Plant-specific sequence fingerprints cause the superhelical arrangement.** **a**, Ribbon diagram of the convex side of LRRs 9–25. The non-canonical  $\beta$ -strands and the Ile-Pro spine are shown in dark and light blue, respectively. **b**, Top view of the BRI1 ectodomain. Disulphide bridges are in

yellow, the N- and C-terminal caps are in pink. **c**, Sequence alignment of LRRs in BRI1, other plant receptor kinases<sup>3–5,20,21</sup> and PGIP<sup>30</sup>. At, *Arabidopsis thaliana*. Pv, *Phaseolus vulgaris*. The canonical LRR consensus sequences are highlighted in blue, plant-specific motifs in green.

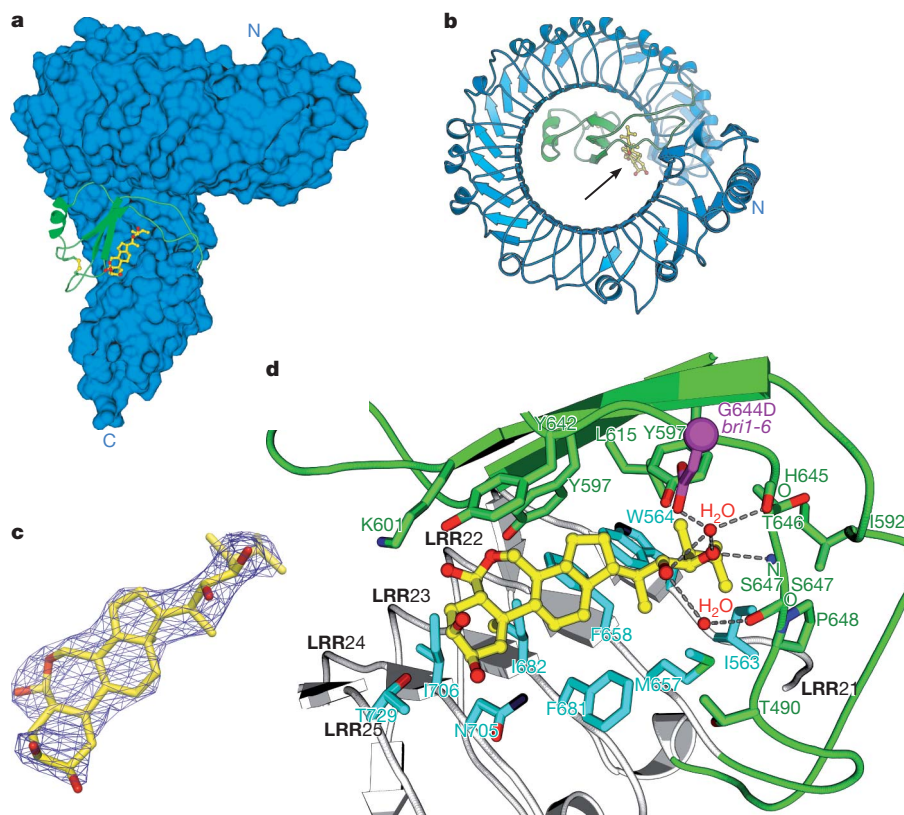
## The island domain

The island domain in BRI1 corresponds to a large insertion in the regular repeat-structure between LRRs 21 and 22 (residues 584–654; Fig. 1a). The resulting ~70-residue segment forms a small domain that folds back into the interior of the superhelix, where it makes extensive polar and hydrophobic interactions with LRRs 13–25 (Fig. 1a, Supplementary Fig. 6 and Supplementary Table 2). The domain fold is characterized by an anti-parallel  $\beta$ -sheet, which is sandwiched between the LRR core and a  $3_{10}$  helix and stabilized by a disulphide bridge (Fig. 3a, Supplementary Fig. 4). The loss-of-function alleles *bri1-9* (Ser662Phe, weak)<sup>32</sup> and *bri1-113* (Gly611Glu, strong)<sup>2</sup> map to this island domain–LRR interface (Supplementary Fig. 6), and probably interfere with folding of the island domain<sup>33</sup>. Two long loops that connect the island domain to the LRR core appear partially disordered in the unliganded receptor (Supplementary Fig. 7). The insertion of a folded domain into the LRR repeat has not been observed in other LRR receptor structures, and is probably an adaptation to the challenge of sensing a small steroid ligand (rather than larger ligands, such as proteins, nucleic acids, or lipids<sup>25,26</sup>).

We next solved a 2.5 Å co-crystal structure with brassinolide, a potent *Arabidopsis* steroid that binds BRI1 with nanomolar affinity<sup>7,34</sup>. One molecule of brassinolide per BRI1 monomer binds in close proximity to the island domain (Fig. 3a–c), which was previously implicated in steroid binding<sup>7,16</sup>. Our structure reveals that the LRR superhelix and the island domain both extensively contribute to formation of the hormone binding site. The A–D rings of the steroid bind to a hydrophobic surface, which is provided by LRRs 23–25 and that maps to the inner side of the BRI1 superhelix (Fig. 3b, d, Supplementary Fig. 8). The alkyl chain of the hormone fits into a small pocket formed by residues originating from LRRs 21 and 22 (Ile 563, Trp 564, Met 657, Phe 658)

and from two loops connecting the island domain with the LRR core (Fig. 3d). The hydrophobic nature and restricted size of this pocket now explain why steroid ligands with bulkier or charged alkyl side chains, such as the arthropod steroid ecdysone (Supplementary Fig. 8), cannot be recognized by BRI1 (ref. 7). A few polar interactions with brassinolide's second diol moiety (Fig. 3d) are established with Tyr 597 and main-chain atoms from His 645 and Ser 647 in the island domain, and are mediated by water molecules (Fig. 3d). Mutation of the neighbouring Gly 644 to Asp may interfere with this hydrogen-bonding network, and explain why this mutation greatly reduces the binding activity of the receptor<sup>7</sup> and causes the loss-of-function phenotype *bri1-6* (ref. 32; Fig. 3d). No polar contacts are observed with the seven-membered B-ring lactone (Fig. 3d), consistent with B-ring modifications as found in, for example, castasterone (Supplementary Fig. 8) being tolerated by BRI1 (refs 7, 35).

The steroid-complex reveals a hormone-binding site that involves a much larger portion of the LRR domain than previously anticipated<sup>16</sup>. Major interactions between the steroid and the BRI1 ectodomain originate from the very C-terminal LRRs 23–25, which bring the hormone in close proximity to the membrane (Fig. 3a, d). Importantly, while there is a significant hormone–receptor interface (550 Å<sup>2</sup>) for such a small molecule ligand, large parts of the steroid are exposed to the solvent, including the 2 $\alpha$ ,3 $\alpha$ -diol moiety in brassinolide that is important for biological activity<sup>35</sup>. Thus, protein–protein interactions may be involved in the recognition of the steroid ligand, with the hormone itself providing a docking platform. Importantly, steroid binding induces a conformational rearrangement and fixing of the island domain, which becomes fully ordered and competent to participate in protein–protein interactions that could be critical for receptor activation (Supplementary Fig. 7).



**Figure 3 | The steroid hormone binding site maps to the C-terminal inner surface of the superhelix.** **a**, Brassinolide (yellow sticks) binds to a surface provided by the LRR domain (in blue) and by parts of the island domain (green ribbon). **b**, Location of the steroid (arrowed) in the centre of the BRI1 superhelix. **c**, Close-up view of brassinolide, including an omit  $2F_o - F_c$  electron density map contoured at 1.5  $\sigma$ . **d**, Protein–hormone interactions in the BRI1 steroid binding site. Ribbon diagram of LRRs 21–25 (in grey) is shown, together with parts of the island domain (in green). Contacting residues are in full side-chain representation, polar interactions are dotted lines, and water molecules are red spheres. *bri1-6* (Gly644Asp) is depicted in magenta.

electron density map contoured at 1.5  $\sigma$ . **d**, Protein–hormone interactions in the BRI1 steroid binding site. Ribbon diagram of LRRs 21–25 (in grey) is shown, together with parts of the island domain (in green). Contacting residues are in full side-chain representation, polar interactions are dotted lines, and water molecules are red spheres. *bri1-6* (Gly644Asp) is depicted in magenta.



## A protein interaction platform

Four known BRI1 mis-sense alleles map to the inner surface of the last five LRRs (Fig. 4a). This surface is not masked by carbohydrate (Supplementary Fig. 9), and contains both the hormone-binding site and the island domain (Figs 3a, d and 4a). Three mutations cluster in a loop connecting the island domain with LRR 22 (Fig. 4a). This loop is partially disordered in the unliganded structure but is well-defined in the brassinolide complex (Supplementary Fig. 7). We speculate that this loop, when ordered, is engaged in protein–protein interactions that are critical for receptor activation, and that mis-sense alleles in BRI1 modulate these interactions. The gain-of-function allele *sud1* (ref. 36; Gly643Glu) may establish contact with Ser 623 in the island domain, and lead to an ordered loop even in the absence of steroid ligand (Supplementary Fig. 10). Mutation of the neighbouring Gly 644 to Asp causes the loss-of-function phenotype *bri1-6* (ref. 32; see above, and Figs 3d, 4a), and mutation of conserved Thr 649 to Lys inactivates barley BRI1 (ref. 37). These mutations, when modelled *in silico*, induce steric clashes with residues in the island domain and in the underlying LRR domain (Supplementary Fig. 10), and thus may distort the position of the loop. Interestingly, *bri1-102*, a strong loss-of-function mutation (Thr750Ile)<sup>38</sup> that does not affect steroid binding<sup>7</sup>, maps to a distinct surface area in LRR 25 (Fig. 4a). Thus protein–protein interactions critical for receptor activation may not be restricted to the island domain, but also involve residues from the LRR core.

## Receptor activation

BRI1 has been shown to exist at least partially as a homo-oligomer *in planta*<sup>13,39,40</sup>. Thus steroid binding to the island domain and the concomitant rearrangements of the island domain loop could induce a conformational change in a preformed BRI1 homodimer<sup>13</sup>, or allow for ligand-dependent dimerization of the BRI1 ectodomain. However, models of BRI1 dimers that bring the C termini of their ectodomains into close proximity (we note that the cytoplasmic kinase domains of BRI1 can interact<sup>13,18</sup>) and that make use of the interaction surface outlined above, encounter steric clashes with the N-terminal LRRs (Supplementary Fig. 11). Furthermore, in contrast to TLR ectodomains, which in crystals tend to form homodimers even in the absence of ligand<sup>24,25</sup>, dimers cannot be seen in BRI1 crystals grown under the same acidic pH conditions that are typically associated with the plant cell wall. The largest interface area between two neighbouring BRI1 molecules amounts to only ~1.5% of the total accessible surface area, consistent with the high solvent content of our crystals (see Supplementary Methods). The main crystal contact involves a head-to-head arrangement of two BRI1 monomers, a configuration that would

place the cytoplasmic kinase domains far apart (Supplementary Fig. 12). In size-exclusion chromatography experiments, the recombinant BRI1 ectodomain elutes as a monomer in the absence of steroid ligand, and shows no tendency to dimerize or oligomerize in the presence of a ~4 times molar excess of brassinolide (Fig. 4b).

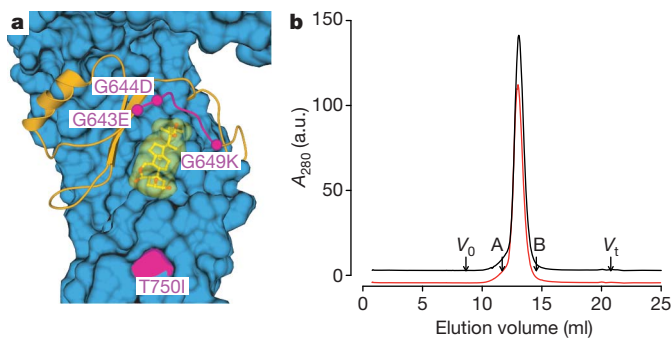
Our analyses suggest that the superhelical BRI1 LRR domain alone has no tendency to oligomerize, indicating that BRI1 receptor activation may not be mediated by ligand-induced homodimerization of the ectodomain (as described for TLRs<sup>25,26</sup>) or by conformational changes in preformed homodimers<sup>13</sup>. We do not dismiss the facts that the cytoplasmic kinase domain of BRI1 can dimerize<sup>18</sup>, or that BRI1 homo-oligomers are present *in vivo*<sup>13,39,40</sup>. However, our structures reinforce the notion that homo-oligomerization of BRI1 may be constitutive and independent of ligand stimulus<sup>39</sup>. The presence of an interaction platform that undergoes conformational changes when steroid binds and that harbours several loss- and gain-of-function alleles suggests that interaction with another protein factor may control BRI1 activation.

## Discussion

The structure of the BRI1 ectodomain offers several new insights, and its twisted shape is likely to characterize the architecture of many plant LRR-RKs<sup>1</sup>. The presence of a folded-domain insert appears to be an adaptation to the recognition of a small molecule ligand, a challenge that smaller LRR proteins have met by generating loop insertions into their capping motifs<sup>41</sup>. BRI1's fascinating mode of ligand recognition reveals how steroids can be sensed at membranes and rationalizes a large set of genetic and biochemical findings<sup>2,7,32,38</sup>.

Different BRI1 receptor activation mechanisms have been proposed, including ligand-dependent dimerization as seen for TLRs<sup>25,26</sup> and ligand-induced conformational changes in preformed homodimers<sup>13</sup>. Our analyses suggest that the superhelical shape of the BRI1 ectodomain is incompatible with homodimerization, and that the isolated ectodomain behaves as a monomer even in the presence of steroid. These findings leave us with the alternative hypothesis that another protein factor could bind to the interaction platform in BRI1 that would minimally encompass the steroid ligand, LRRs 21–25 and parts of the island domain (Fig. 4a). Although it is possible that an unknown protein fulfils this role and provides a dimerization interface for two BRI1 molecules (as seen, for example, for TLR4; ref. 26), genetic and biochemical screens have not uncovered this protein. It is thus possible that the small receptor kinase BAK1 acts as a direct brassinosteroid co-receptor, as suggested previously<sup>10,20</sup>. It has been demonstrated that BAK1 is a genetic component of the brassinosteroid pathway<sup>20,21</sup>, that BRI1 and BAK1 interact in a steroid-dependent manner<sup>22</sup> and that both receptors trans-phosphorylate each other on ligand stimulus<sup>22</sup>. Notably, a homology model of the BAK1 ectodomain (Supplementary Fig. 13) is compatible in size and shape with the interaction platform in BRI1, and the BAK1 *elg* allele, which maps to the BAK1 ectodomain (Supplementary Fig. 14), renders plants hypersensitive to brassinosteroid treatment<sup>42</sup>. We speculate that the *sud1*, *bri1-6*, *bri1-102* and *elg* mutations modulate the interaction between the BRI1 and BAK1 ectodomains in a brassinosteroid-dependent manner (Supplementary Fig. 14). The demonstration that BAK1 is essential for brassinosteroid sensing may have been obscured owing to genetic redundancy<sup>20</sup>, with at least two BAK1-like proteins interacting with BRI1 *in vivo*<sup>43,44</sup>. We recently overcame this limitation by showing that the BRI1 inhibitor protein BKI1 blocks the interaction between the BAK1 and BRI1 kinase domains<sup>18</sup>. Importantly, transgenic lines that constitutively deliver BKI1 to the site of BRI1 signalling resemble strong BRI1 loss-of-function mutants, suggesting an important role for receptor–co-receptor association in brassinosteroid signal initiation<sup>18</sup>.

Future studies will undoubtedly test this heteromerization model and dissect the relative contributions of the BRI1 and BAK1 ectodomains, their transmembrane segments and their cytoplasmic kinase domains to receptor activation. It will become important then to



**Figure 4 | An accessible membrane-proximal region of BRI1 may provide a protein–protein interaction platform.** **a**, Overview of the C-terminal surface area not masked by carbohydrate. Brassinolide is shown in yellow, the island domain in orange, and genetic alleles in magenta. **b**, Analytical gel-filtration trace (absorbance at 280 nm). The free ectodomain elutes as a monomer (black line), as does a putative complex with brassinolide (red line). Void volume ( $V_0$ ) and total volume ( $V_t$ ) are shown, together with elution volumes for molecular mass standards (A, aldolase, molecular mass 158,000 Da; B, conalbumin, molecular mass 75,000 Da). The calculated molecular mass for the monomer peak is ~125 kDa. The molecular mass of purified BRI1 is ~110 kDa.

understand how BAK1 could serve as co-receptor for other LRR-RK signalling pathways<sup>5,19</sup>.

## METHODS SUMMARY

The BRI1 ectodomain (residues 29–788) was produced by secreted expression in baculovirus-infected insect cells, harvested 4 d after infection by ultrafiltration and purified by tandem-affinity chromatography, then by gel filtration. BRI1 was concentrated to 15 mg ml<sup>-1</sup> and crystallized by vapour diffusion using a reservoir solution containing 14% PEG 4,000, 0.2 M (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>, 0.1 M citric acid (pH 4.0). The brassinolide complex was obtained by co-crystallization. Diffraction data to 2.5 Å resolution were collected on a rotating anode X-ray generator and at beamline 8.2.1 of the Advanced Light Source (ALS), Berkeley. The structure was solved using the single isomorphous replacement method. Data and refinement statistics are summarized in Supplementary Table 1.

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

Received 21 January; accepted 26 April 2011.

Published online 12 June 2011.

- Shiu, S. H. & Bleeker, A. B. Receptor-like kinases from *Arabidopsis* form a monophyletic gene family related to animal receptor kinases. *Proc. Natl Acad. Sci. USA* **98**, 10763–10768 (2001).
- Li, J. & Chory, J. A putative leucine-rich repeat receptor kinase involved in brassinosteroid signal transduction. *Cell* **90**, 929–938 (1997).
- Clark, S. E., Williams, R. W. & Meyerowitz, E. M. The *CLAVATA1* gene encodes a putative receptor kinase that controls shoot and floral meristem size in *Arabidopsis*. *Cell* **89**, 575–585 (1997).
- Nadeau, J. A. & Sack, F. D. Control of stomatal distribution on the *Arabidopsis* leaf surface. *Science* **296**, 1697–1700 (2002).
- Gómez-Gómez, L. & Boller, T. FLS2: an LRR receptor-like kinase involved in the perception of the bacterial elicitor flagellin in *Arabidopsis*. *Mol. Cell* **5**, 1003–1011 (2000).
- Nishimura, R. *et al.* HAR1 mediates systemic regulation of symbiotic organ development. *Nature* **420**, 426–429 (2002).
- Wang, Z. Y., Seto, H., Fujioka, S., Yoshida, S. & Chory, J. BRI1 is a critical component of a plasma-membrane receptor for plant steroids. *Nature* **410**, 380–383 (2001).
- Ogawa, M., Shinohara, H., Sakagami, Y. & Matsubayashi, Y. *Arabidopsis* CLV3 peptide directly binds CLV1 ectodomain. *Science* **319**, 294 (2008).
- Belkhadir, Y. & Chory, J. Brassinosteroid signaling: a paradigm for steroid hormone signaling from the cell surface. *Science* **314**, 1410–1411 (2006).
- Vert, G., Nemhauser, J. L., Geldner, N., Hong, F. & Chory, J. Molecular mechanisms of steroid hormone signaling in plants. *Annu. Rev. Cell Dev. Biol.* **21**, 177–201 (2005).
- Mangelsdorf, D. J. *et al.* The nuclear receptor superfamily: the second decade. *Cell* **83**, 835–839 (1995).
- Geldner, N., Hyman, D. L., Wang, X., Schumacher, K. & Chory, J. Endosomal signaling of plant steroid receptor kinase BRI1. *Genes Dev.* **21**, 1598–1602 (2007).
- Wang, X. *et al.* Autoregulation and homodimerization are involved in the activation of the plant steroid receptor BRI1. *Dev. Cell* **8**, 855–865 (2005).
- Wang, X. & Chory, J. Brassinosteroids regulate dissociation of BK1, a negative regulator of BRI1 signaling, from the plasma membrane. *Science* **313**, 1118–1122 (2006).
- He, Z. *et al.* Perception of brassinosteroids by the extracellular domain of the receptor kinase BRI1. *Science* **288**, 2360–2363 (2000).
- Kinoshita, T. *et al.* Binding of brassinosteroids to the extracellular domain of plant receptor kinase BRI1. *Nature* **433**, 167–171 (2005).
- Wang, X. *et al.* Identification and functional analysis of *in vivo* phosphorylation sites of the *Arabidopsis* BRASSINOSTEROID-INSENSITIVE1 receptor kinase. *Plant Cell* **17**, 1685–1703 (2005).
- Jailais, Y. *et al.* Tyrosine phosphorylation controls brassinosteroid receptor activation by triggering membrane release of its kinase inhibitor. *Genes Dev.* **25**, 232–237 (2011).
- Chinchilla, D., Shan, L., He, P., de Vries, S. & Kemmerling, B. One for all: the receptor-associated kinase BAK1. *Trends Plant Sci.* **14**, 535–541 (2009).
- Nam, K. H. & Li, J. BRI1/BAK1, a receptor kinase pair mediating brassinosteroid signaling. *Cell* **110**, 203–212 (2002).
- Li, J. *et al.* BAK1, an *Arabidopsis* LRR receptor-like protein kinase, interacts with BRI1 and modulates brassinosteroid signaling. *Cell* **110**, 213–222 (2002).
- Wang, X. *et al.* Sequential transphosphorylation of the BRI1/BAK1 receptor kinase complex impacts early events in brassinosteroid signaling. *Dev. Cell* **15**, 220–235 (2008).
- Kim, T.-W. & Wang, Z.-Y. Brassinosteroid signal transduction from receptor kinases to transcription factors. *Annu. Rev. Plant Biol.* **61**, 681–704 (2010).
- Choe, J., Kelker, M. S. & Wilson, I. A. Crystal structure of human toll-like receptor 3 (TLR3) ectodomain. *Science* **309**, 581–585 (2005).
- Liu, L. *et al.* Structural basis of toll-like receptor 3 signaling with double-stranded RNA. *Science* **320**, 379–381 (2008).
- Park, B. S. *et al.* The structural basis of lipopolysaccharide recognition by the TLR4-MD-2 complex. *Nature* **458**, 1191–1195 (2009).
- Bella, J., Hindle, K. L., McEwan, P. A. & Lovell, S. C. The leucine-rich repeat structure. *Cell. Mol. Life Sci.* **65**, 2307–2333 (2008).
- Evdokimov, A. G., Anderson, D. E., Routzahn, K. M. & Waugh, D. S. Unusual molecular architecture of the *Yersinia pestis* cytotoxin YopM: a leucine-rich repeat protein with the shortest repeating unit. *J. Mol. Biol.* **312**, 807–821 (2001).
- Schubert, W. D. *et al.* Structure of internalin, a major invasion protein of *Listeria monocytogenes*, in complex with its human receptor E-cadherin. *Cell* **111**, 825–836 (2002).
- Di Matteo, A. *et al.* The crystal structure of polygalacturonase-inhibiting protein (PGIP), a leucine-rich repeat protein involved in plant defense. *Proc. Natl Acad. Sci. USA* **100**, 10124–10128 (2003).
- Kajava, A. V. Structural diversity of leucine-rich repeat proteins. *J. Mol. Biol.* **277**, 519–527 (1998).
- Noguchi, T. *et al.* Brassinosteroid-insensitive dwarf mutants of *Arabidopsis* accumulate brassinosteroids. *Plant Physiol.* **121**, 743–752 (1999).
- Jin, H., Yan, Z., Nam, K. H. & Li, J. Allele-specific suppression of a defective brassinosteroid receptor reveals a physiological role of UGGT in ER quality control. *Mol. Cell* **26**, 821–830 (2007).
- Fujioka, S. *et al.* The *Arabidopsis* deetiolated2 mutant is blocked early in brassinosteroid biosynthesis. *Plant Cell* **9**, 1951–1962 (1997).
- Back, T. G. & Pharis, R. P. Structure-activity studies of brassinosteroids and the search for novel analogues and mimetics with improved bioactivity. *J. Plant Growth Regul.* **22**, 350–361 (2003).
- Diévar, A., Hymes, M. J., Li, J. & Clark, S. E. Brassinosteroid-independent function of BRI1/CLV1 chimeric receptors. *Funct. Plant Biol.* **33**, 723–730 (2006).
- Gruszka, D., Szarejko, I. & Maluszynski, M. New allele of *HvBRI1* gene encoding brassinosteroid receptor in barley. *J. Appl. Genet.* (published online, doi:10.1007/s13353-011-0031-7 (8 February 2011)).
- Friedrichsen, D. M., Joazeiro, C. A., Li, J., Hunter, T. & Chory, J. Brassinosteroid-insensitive-1 is a ubiquitously expressed leucine-rich repeat receptor serine/threonine kinase. *Plant Physiol.* **123**, 1247–1256 (2000).
- Hink, M. A., Shah, K., Russinova, E., de Vries, S. C. & Visser, A. J. W. G. Fluorescence fluctuation analysis of *Arabidopsis thaliana* somatic embryogenesis receptor-like kinase and brassinosteroid insensitive 1 receptor oligomerization. *Biophys. J.* **94**, 1052–1062 (2008).
- Russinova, E. *et al.* Heterodimerization and endocytosis of *Arabidopsis* brassinosteroid receptors BRI1 and AtSERK3 (BAK1). *Plant Cell* **16**, 3216–3229 (2004).
- Han, B. W., Herrin, B. R., Cooper, M. D. & Wilson, I. A. Antigen recognition by variable lymphocyte receptors. *Science* **321**, 1834–1837 (2008).
- Whippo, C. W. & Hangarter, R. P. A brassinosteroid-hypersensitive mutant of BAK1 indicates that a convergence of photomorphogenic and hormonal signaling modulates phototropism. *Plant Physiol.* **139**, 448–457 (2005).
- He, K. *et al.* BAK1 and BKK1 regulate brassinosteroid-dependent growth and brassinosteroid-independent cell-death pathways. *Curr. Biol.* **17**, 1109–1115 (2007).
- Karlova, R. *et al.* The *Arabidopsis* SOMATIC EMBRYOGENESIS RECEPTOR-LIKE KINASE1 protein complex includes BRASSINOSTEROID-INSENSITIVE1. *Plant Cell* **18**, 626–638 (2006).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank J. Vanhansy and W. Yu for maintaining insect cell stocks, M. Jinek and B. W. Han for advice, W. Kwiatowski for maintenance of the Salk X-ray equipment, Y. Jailais for discussion, and F. V. Chisari for encouragement and support. This work was supported by the Howard Hughes Medical Institute and a grant from the National Science Foundation (IOS-0649389) to J.C. M.H. was supported by long-term fellowships from the European Molecular Biology Organisation and the International Human Frontier Science Program Organisation. Y.B. was a Howard Hughes Medical Institute fellow of the Life Sciences Research Foundation and also received support from the Philippe Foundation. I.A.W. was supported by NIH grant AI042266 and by the Skaggs Institute for Chemical Biology.

**Author Contributions** M.H., Y.B., J.P.N. and J.C. designed the project. M.H. expressed the BRI1 ectodomain in the laboratory of I.A.W. with initial help from Y.B. M.H. purified and crystallized the protein, and phased and refined the structures. M.D. determined viral titres and optimized production of viruses. T.D. cloned the modified transfer vector. M.H., I.A.W. and J.C. analysed the data. J.C. supervised the project. M.H. wrote the paper with input from the other authors.

**Author Information** Atomic coordinates and structure factors for the reported crystal structures have been deposited in the Protein Data Bank under accession numbers 3RIZ for the unliganded BRI1 ectodomain and 3RJO for the BRI1–brassinolide complex. Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at [www.nature.com/nature](http://www.nature.com/nature). Correspondence and requests for materials should be addressed to J.C. ([chory@salk.edu](mailto:chory@salk.edu)).

## METHODS

**Protein expression and purification.** A synthetic gene comprising the entire BRI1 ectodomain (residues 29–788) and codon optimized for expression in *Trichoplusia ni* was synthesized by Geneart. The gene was cloned into a modified pBAC-6 transfer vector (Novagen), providing a glycoprotein 64 signal peptide and a C-terminal TEV (tobacco etch virus protease) cleavable Strep-9xHis tandem-affinity tag. Recombinant baculoviruses were generated by co-transfecting the transfer vector with linearized baculovirus DNA (ProFold-ER1, AB vector) and amplified in Sf9 cells. The fusion protein was expressed in Hi5 cells using a multiplicity of infection of 5, and harvested from the medium 4 days after infection by tangential flow filtration using a 30 kDa MWCO (molecular weight cut-off) filter membrane (GE Healthcare). BRI1 was purified by sequential Co<sup>2+</sup> (His select gel, Sigma) and Strep (Strep-Tactin Superflow high-capacity, IBA) affinity chromatography. Next, the tandem-affinity tag was removed by incubating purified BRI1 with recombinant TEV protease in 1:100 molar ratio. The cleaved tag and the protease were separated from BRI1 by size-exclusion chromatography on a Superdex 200 HR10/30 column (GE Healthcare) equilibrated in 20 mM HEPES (pH 7.5), 100 mM NaCl, 1 mM EDTA. Monomeric peak fractions were concentrated to ~15 mg ml<sup>-1</sup> and snap frozen in liquid nitrogen. About 50–80 µg of purified BRI1 could be obtained from 1 l of insect cell culture.

**Crystallization and data collection.** Initial crystals of BRI1 appeared in 18% PEG 4,000, 0.8 M KCl using the counter diffusion method. Diffraction quality crystals of about 300 × 80 × 600 µm could be grown after multiple rounds of microseeding at room temperature by vapour diffusion in hanging drops composed of 1.25 µl of protein solution (15 mg ml<sup>-1</sup>) and 1.25 µl of crystallization buffer (14% PEG 4,000, 0.2 M (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>, 0.1 M citric acid pH 4.0) suspended above 1.0 ml of the mother liquor as the reservoir solution. For structure solution crystals were stabilized, derivatized and cryo-protected by serial transfer into 16% PEG 4,000, 1.7 M Na malonate (pH 4.0) and 0.5 M NaI, and cryo-cooled in liquid nitrogen. Single-wavelength anomalous diffraction (SAD) data to 2.9 Å resolution were collected on a Rigaku MicroMax rotating anode equipped with a copper filament ( $\lambda = 1.5418$  Å), Osmic mirrors and an R-Axis IV++ detector. Native crystals were transferred to a cryo-protective solution containing 16% PEG 4,000 and 1.7 M Na malonate (pH 4.0) and flash-cooled in liquid nitrogen. An isomorphous native data set to 2.5 Å was collected at beam-line 8.2.1 ( $\lambda = 0.9998$  Å) of the Advanced Light Source (ALS), Berkeley. The hormone-bound structure was obtained by dissolving brassinolide (Chemical clones Inc.) to a concentration of 1 mM in 100% DMSO. This stock solution was diluted to a final concentration of about 50 µM in protein storage buffer (20 mM HEPES pH 7.5, 100 mM NaCl, 1 mM EDTA). Purified BRI1 protein was added to a final concentration of about 12.5 µM (1.5 mg ml<sup>-1</sup>) and the mixture was incubated at room temperature for 16 h. Next, the complex was re-concentrated to 18 mg ml<sup>-1</sup>, and immediately used for crystallization. Crystals appeared under similar conditions as established for the unbound form and diffracted again to about 2.5 Å ( $\lambda = 1.5418$  Å). Data processing and scaling was done with XDS<sup>45</sup> (version: May 2010) (Supplementary Table 1).

**Structure solution and refinement.** The program XPREP (Bruker AXS) was used to scale native and derivative data for SIRAS (single isomorphous replacement with anomalous scattering) analysis. Using data between 30 and 3.7 Å,

SHELXD<sup>46</sup> located 52 iodine sites (CC All/Weak 42.50/19.82). 16 consistent sites were input into the program SHARP<sup>47</sup> for phasing and identification of 10 additional sites at 2.9 Å resolution (Supplementary Fig. 1a). Refined heavy atom sites and phases were input into phenix.resolve<sup>48</sup> for density modification and phase extension to 2.5 Å (final FOM was 0.55). The resulting electron density map was readily interpretable (Supplementary Fig. 1b), and the structure was completed in alternating cycles of model building in COOT<sup>49</sup> and restrained TLS refinement in phenix.refine (<http://www.phenix-online.org>). Refinement statistics are summarized in Supplementary Table 1. The crystals contain one BRI1 monomer per asymmetric unit with a solvent content of ~60%. The final models comprise residues 29–771, with the C termini (residues 772–788) being completely disordered. The structure contains 25 LRRs as initially proposed<sup>3</sup>, and not 24 LRRs as concluded from later modelling studies<sup>10</sup>. Loop residues 590, 637 and 638 in the island domain appear disordered in the unliganded structure. Amino acids whose side chains could not be modelled with confidence were truncated to alanine (2% of all residues). Analysis with MolProbity<sup>50</sup> suggested that both refined models have excellent stereochemistry, with the free form having 93.3% of all residues in the favoured region of the Ramachandran plot, and no outliers (Molprobity score is 2.2 corresponding to the 90th percentile for structures ( $N = 6,681$ ) at  $2.52 \pm 0.25$  Å resolution). The brassinolide complex structure has 92.7% of all residues in the favoured region of the Ramachandran plot and no outliers (Molprobity score is 2.3 corresponding to the 86th percentile for structures ( $N = 6,632$ ) at  $2.54 \pm 0.25$  Å resolution).

**Size-exclusion chromatography.** This was performed using a Superdex 200 HR 10/30 column (GE Healthcare) pre-equilibrated in 25 mM citric acid/sodium citrate buffer (pH 4.5), 100 mM NaCl. 100 µl of sample (5 mg ml<sup>-1</sup>) was loaded onto the column and elution at 0.6 ml min<sup>-1</sup> was monitored by ultraviolet absorbance at 280 nm. Incubation with brassinolide was performed as described in the crystallization section.

**Homology modelling.** Homology modelling of the AtBAK1 ectodomain (residues 27–227; Uniprot <http://www.uniprot.org>; accession Q94F62) was performed with the program MODELLER (<http://www.salilab.org/modeller/>) using the BRI1 and PGIP<sup>30</sup> (PDB 1OGQ) structures as template. Structure-based sequence alignments were done using T-COFFEE (<http://www.tcoffee.org>). BRI1 and BAK1 share ~35%, PGIP and BAK1 share ~31% sequence identity, with the LRR and N-cap consensus sequences being highly conserved.

45. Kabsch, W. Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. *J. Appl. Crystallogr.* **26**, 795–800 (1993).
46. Sheldrick, G. M. A short history of SHELX. *Acta Crystallogr. A* **64**, 112–122 (2008).
47. Bricogne, G., Vonrhein, C., Flensburg, C., Schiltz, M. & Paoletti, W. Generation, representation and flow of phase information in structure determination: recent developments in and around SHARP 2.0. *Acta Crystallogr. D* **59**, 2023–2030 (2003).
48. Terwilliger, T. C. *et al.* Iterative model building, structure refinement and density modification with the PHENIX AutoBuild wizard. *Acta Crystallogr. D* **64**, 61–69 (2008).
49. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **60**, 2126–2132 (2004).
50. Davis, I. W. *et al.* MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res.* **35**, W375–W383 (2007).



# Structure-based design of non-natural amino-acid inhibitors of amyloid fibril formation

Stuart A. Sievers<sup>1\*</sup>, John Karanicolas<sup>2,3\*</sup>, Howard W. Chang<sup>1\*</sup>, Anni Zhao<sup>1\*</sup>, Lin Jiang<sup>1\*</sup>, Onofrio Zirafi<sup>4</sup>, Jason T. Stevens<sup>3</sup>, Jan Münch<sup>4</sup>, David Baker<sup>2</sup> & David Eisenberg<sup>1</sup>

Many globular and natively disordered proteins can convert into amyloid fibrils. These fibrils are associated with numerous pathologies<sup>1</sup> as well as with normal cellular functions<sup>2,3</sup>, and frequently form during protein denaturation<sup>4,5</sup>. Inhibitors of pathological amyloid fibril formation could be useful in the development of therapeutics, provided that the inhibitors were specific enough to avoid interfering with normal processes. Here we show that computer-aided, structure-based design can yield highly specific peptide inhibitors of amyloid formation. Using known atomic structures of segments of amyloid fibrils as templates, we have designed and characterized an all-D-amino-acid inhibitor of the fibril formation of the tau protein associated with Alzheimer's disease, and a non-natural L-amino-acid inhibitor of an amyloid fibril that enhances sexual transmission of human immunodeficiency virus. Our results indicate that peptides from structure-based designs can disrupt the fibril formation of full-length proteins, including those, such as tau protein, that lack fully ordered native structures. Because the inhibiting peptides have been designed on structures of dual- $\beta$ -sheet 'steric zippers', the successful inhibition of amyloid fibril formation strengthens the hypothesis that amyloid spines contain steric zippers.

The finding that dozens of pathologies, including Alzheimer's disease, are associated with amyloid fibrils has stimulated research on fibril inhibition. One approach uses the self-associating property of proteins that form fibrils to poison fibril formation with short peptide segments<sup>6–11</sup>. A second approach is based on screening for molecules that can disrupt fibril formation<sup>12,13</sup>. Here we take a third approach to fibril inhibition: structure-based design of non-natural peptides targeted to block the ends of fibrils. With advanced sampling techniques and by minimizing an appropriate energy function, we identify novel candidate inhibitors computationally from a large peptide space that interact favourably with our template structure. This approach has been made possible by the determination of several dozen fibril-like atomic structures of segments from amyloid-forming proteins<sup>14–16</sup>.

These structures reveal a common motif called a steric zipper, in which a pair of  $\beta$ -sheets is held together by the interdigitation of their side chains<sup>14</sup>. Using as templates the steric-zipper structures formed by segments of two pathological proteins, we have designed inhibitors that cap fibril ends. As we show, the inhibitors greatly slow the fibril formation of the parent proteins of the segments, offering a route to designed chemical interventions and supporting the hypothesis that steric zippers are the principal structural elements of these fibrils.

One of the two fibril-like steric zippers that we have chosen as a target for inhibitor design is the hexapeptide VQIVYK, residues 306–311 of the tau protein, which forms intracellular amyloid fibrils in Alzheimer's disease<sup>17</sup>. This segment has been shown to be important for fibril formation of the full-length protein and itself forms fibrils

with biophysical properties similar to full-length tau fibrils<sup>15,18,19</sup>. Our second template for inhibitor design, identified by the '3D profile' algorithm<sup>20,21</sup>, is the steric-zipper structure of the peptide segment GGVLVN from the amyloid fibril formed by <sup>248</sup>PAP<sup>286</sup>, a proteolytic fragment containing residues 248–286 of prostatic acid phosphatase, a protein abundant in semen. <sup>248</sup>PAP<sup>286</sup> fibrils, also known as semen-derived enhancer of virus infection (SEVI), enhance human immunodeficiency virus (HIV) infection by orders of magnitude in cell culture studies, whereas the monomeric peptide is inactive<sup>22</sup>.

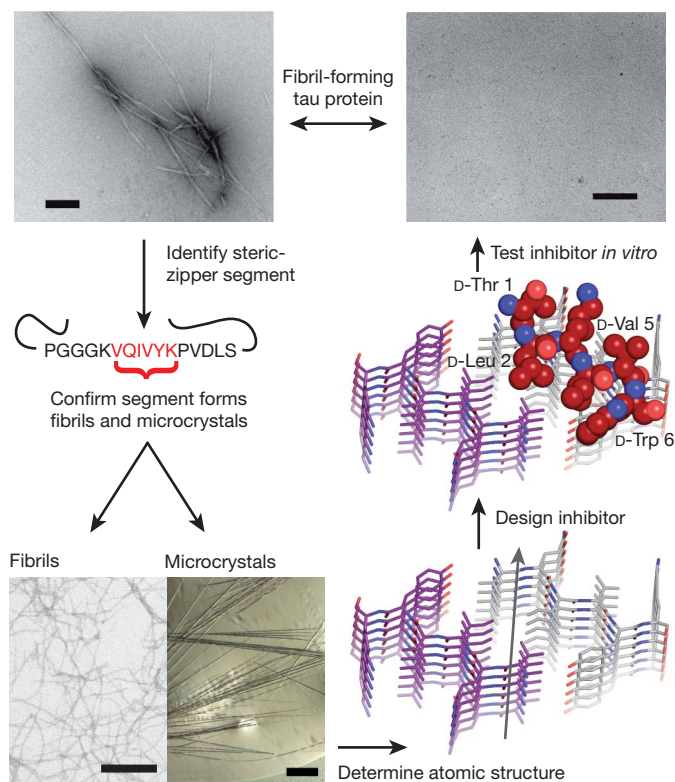
Our computational approach to designing non-natural peptides that inhibit fibril formation is summarized in Fig. 1 for the VQIVYK segment of tau protein; the same general strategy is used for the GGVLVN segment of <sup>248</sup>PAP<sup>286</sup>. In both systems, we design a tight interface between the inhibiting peptide and the end of the steric zipper to block additional segments from joining the fibril. By sampling L or D amino acids, or commercially available non-natural amino acids, we can design candidate inhibitors with side chains that maximize hydrogen bonding and hydrophobic interactions across the interface.

We propose that the steric-zipper structures of the VQIVYK and GGVLVN segments represent the spines of the fibrils formed by the parent proteins containing these segments. Supporting our hypothesis are our results that D-amino-acid inhibitors designed on the VQIVYK steric-zipper template inhibit fibril formation not only of the VQIVYK segment, but also of two tau constructs, K12 and K19<sup>23,24</sup> (Fig. 2a). Similarly, the peptide composed of non-natural amino acids designed on the GGVLVN template inhibits the fibril formation of <sup>248</sup>PAP<sup>286</sup> and greatly inhibits the HIV infectivity of human cells in culture.

To design a D-amino-acid hexapeptide sequence that interacts favourably with the VQIVYK steric zipper<sup>15</sup>, and prevents further addition of tau molecules to the fibril, we used the Rosetta software<sup>25</sup>. This led to the identification of four D-amino-acid peptides: D-TLKIVW, D-TWKLVL, D-DYYFEF and D-YVIIER, in which the prefix signifies that all  $\alpha$ -carbon atoms are in the D configuration (Fig. 2b, c, Supplementary Figs 1 and 2 and Supplementary Table 1). In the D-TLKIVW design model (Fig. 2b, c and Supplementary Fig. 1), the inhibitor packs tightly across the top of the VQIVYK steric-zipper structure, maintaining all main-chain hydrogen bonds. The side-chain hydrogen bonding between layers of stacked Gln 307 residues is replaced in the designed interface by an interaction with D-Lys 3. Several hydrophobic interactions between D-TLKIVW and the two VQIVYK  $\beta$ -strands contribute to the favourable binding energy (Supplementary Table 1). In the design, the D-peptide blocks the addition of another layer of VQIVYK, both above the D-peptide and across on the mating  $\beta$ -sheet (Supplementary Fig. 3). D-Leu 2 of the designed inhibitor prevents the addition of a VQIVYK molecule above it through a steric clash with Ile 308 of VQIVYK and on the mating sheet through a clash with Val 306 and Ile 308 (Supplementary Fig. 3). These steric clashes involving D-Leu 2 are intended to block fibril growth.

<sup>1</sup>Departments of Biological Chemistry and Chemistry and Biochemistry, Howard Hughes Medical Institute, UCLA, Box 951970, Los Angeles, California 90095-1570, USA. <sup>2</sup>Department of Biochemistry and Howard Hughes Medical Institute, University of Washington, Seattle, Washington 98195, USA. <sup>3</sup>Center for Bioinformatics and Department of Molecular Biosciences, University of Kansas, 1200 Sunnyside Avenue, Lawrence, Kansas 66045-7534, USA. <sup>4</sup>Institute of Molecular Virology, University Hospital Ulm, Meyerhofstrasse 1, 89081 Ulm, Germany.

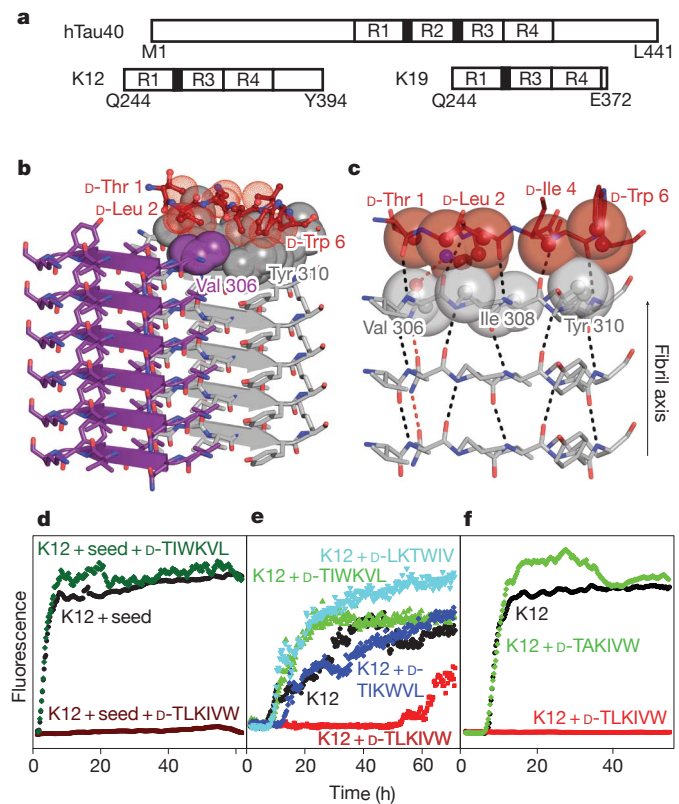
\*These authors contributed equally to this work.



**Figure 1 | Design and characterization of peptide inhibitors of amyloid fibril formation.** Tau constructs form fibrils *in vitro*<sup>24</sup> (top left; scale bar, 200 nm). The VQIVYK segment in isolation forms fibrils and microcrystals (bottom left; fibril scale bar, 200 nm; microcrystal scale bar, 100  $\mu$ m). The atomic structure of the fibril-like VQIVYK segment reveals a characteristic steric-zipper motif<sup>5</sup> comprising a pair of interacting  $\beta$ -sheets (purple and grey) running along the fibril axis (grey arrow) (bottom right). We designed a D-amino-acid peptide to bind to the end of the steric-zipper template and prevent fibril elongation (middle right). The D-peptide (red) is designed to satisfy hydrogen bonds and make favourable non-polar interactions with the molecule below, while preventing the addition of other molecules above and on the opposite  $\beta$ -sheet. As shown *in vitro*, the designed D-peptide prevents the formation of fibrils when incubated with tau K19 (upper right; scale bar, 200 nm).

We used fluorescence spectroscopy and electron microscopy to assess whether the designed D-peptides inhibit the fibril formation of the tau segment VQIVYK and of the tau constructs K12 and K19. Of our designed inhibitors, D-TLKIVW is the most effective (Supplementary Fig. 4). Electron microscopy, performed after three days, verified that incubation with equimolar D-TLKIVW prevents K19 fibril formation, which would otherwise have occurred within the elapsed time (Fig. 1, upper right). D-TLKIVW delays fibril formation of VQIVYK, K12 and K19 even when present in sub-equimolar concentration (Supplementary Fig. 5). A fivefold molar excess of D-TLKIVW delays K12 fibril formation for more than two weeks in some experimental replicates (Supplementary Fig. 5c, d). In tenfold molar excess, D-TLKIVW prevents the fibril formation of K12 for more than 60 hours in the presence of preformed K12 fibril seeds, suggesting that the peptide interacts with fibrils (Fig. 2d). Also, kinetic analysis shows that the fibril elongation rate decreases in the presence of increasing concentrations of inhibitor peptide (Supplementary Fig. 6). The large increase in lag time in unseeded reactions may be due to interactions with small aggregates formed during the process of fibril formation.

To investigate the specificity of the designed inhibitor, we tested scrambled sequence variants of D-TLKIVW that have poor (that is, high) calculated energies and unfavourable packing (Supplementary Table 1). The scrambled peptides D-TIKWVL, D-TIWKVL and D-LKTWIV have little inhibitory effect when present at an equimolar



**Figure 2 | Designed D-peptide delays tau K12 fibril formation in a sequence-specific manner.** a, Tau construct composition<sup>23</sup>. The longest human tau isoform found in the central nervous system, hTau40 (Uniprot ID, P10636-8), contains four microtubule-binding repeats, R1 to R4, whereas K12 and K19 lack R2. The black bars at the amino termini of R2 and R3 represent the fibrillogenic segments VQIINK and VQIVYK, respectively. b, The inhibitor D-TLKIVW (red) is designed to interact with atoms on both  $\beta$ -strands of the VQIVYK steric zipper (grey) primarily through hydrophobic packing and hydrogen-bonding interactions. c, The inhibitor interacts with the VQIVYK  $\beta$ -strand below. The transparent spheres show where the two molecules interact favourably. Black and red dashes indicate main-chain and side-chain hydrogen bonds, respectively. Stereo views of b and c are shown in Supplementary Fig. 1. d, The seeded fibril formation of 50  $\mu$ M K12 in the presence and absence of a tenfold molar excess of peptide was monitored by Thioflavin S fluorescence. In the presence of the scrambled peptide D-TIWKVL (dark green) and alone (black), seeded K12 fibril formation occurs with almost no lag time. However, D-TLKIVW prevents fibril formation for days (maroon). e, At equimolar concentrations, D-TLKIVW (red) inhibits the fibril formation of 50  $\mu$ M K12. D-TIKWVL (blue), with only three residues scrambled, shows weak inhibition. However, no inhibition is observed for either D-TIWKVL (green) or D-LKTWIV (cyan). f, The replacement of D-Leu 2, designed to clash with VQIVYK on the opposite sheet, with D-Ala eliminates the inhibition of fibril formation.

ratio with VQIVYK, K12 and K19 (Fig. 2e and Supplementary Fig. 7), showing that the inhibition is sequence specific. Also, the diastereomer, L-TLKIVW, is less effective than D-TLKIVW (Supplementary Fig. 8). As a further test of the specificity of our design, we confirmed that D-TLKIVW is unable to block the fibril formation of amyloid- $\beta$ , which also is associated with Alzheimer's disease (Supplementary Fig. 9). This suggests that the D-peptide inhibitor is not general to amyloid systems, but is specific to the VQIVYK interface in tau protein. Such specificity is essential for designed inhibitors if they are not to interfere with proteins that natively function in an amyloid state<sup>3</sup>.

To confirm that the designed D-peptide inhibits in accordance with the design model (Fig. 2b, c and Supplementary Fig. 1), we performed several additional tests. First we visualized the position of the inhibitor D-TLKIVW relative to fibrils of the tau construct K19 using electron microscopy. We covalently linked Monomaleimido Nanogold particles



both to the inhibitor and, separately, to a scrambled hexapeptide, D-LKTWIV. We used a blind counting assay and found that, relative to Nanogold alone, D-TLKIVW shows a significant binding preference for the end of fibrils, in contrast to the scrambled control peptide, D-LKTWIV (Fig. 3a and Supplementary Fig. 10).

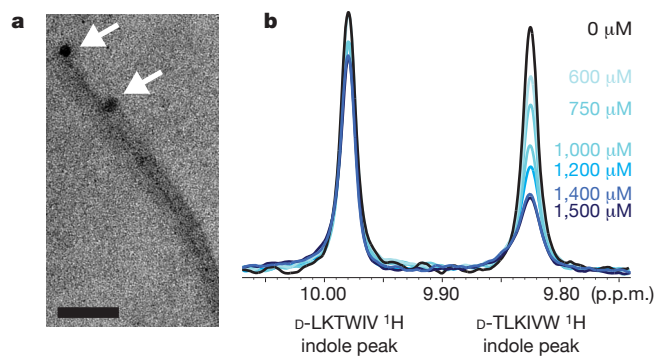
As a further test of the model, we used NMR to characterize the binding affinity of D-TLKIVW for tau fibrils. The  $^1\text{H}$  NMR spectra for D-TLKIVW were collected in the presence of increasing concentrations of VQIVYK or K19 fibrils. Because neither K19 nor VQIVYK contains tryptophan, we were able to monitor the  $^1\text{H}$  resonance of the indole proton of the tryptophan in our inhibitor. When bound to a fibril, the inhibitor, D-TLKIVW, is removed from the soluble phase and the  $^1\text{H}$  resonance is diminished<sup>26</sup> (Fig. 3b and Supplementary Fig. 11). As a control, we also measured spectra for the non-inhibiting peptide D-LKTWIV present with D-TLKIVW in the same reaction mixture. As shown in Fig. 3b, the presence of VQIVYK fibrils at a given concentration reduces the D-TLKIVW indole resonance much more than it does the D-LKTWIV indole resonance. Spectra of the two peptides are shown in Supplementary Fig. 12. By monitoring the D-TLKIVW indole resonance over a range of VQIVYK fibril concentrations, we estimate the apparent dissociation constant of the interaction between D-TLKIVW and VQIVYK fibrils to be  $\sim 2\ \mu\text{M}$  (Supplementary Fig. 11a and Methods). This value corresponds to a standard free binding energy of  $\sim 7.4\ \text{kcal mol}^{-1}$ , with  $\sim 2.5\ \text{kcal mol}^{-1}$  from non-polar interactions and  $\sim 4.9\ \text{kcal mol}^{-1}$  from six hydrogen bonds (Methods). Repeating the NMR binding experiment with K19 fibrils yields a similar trend (Supplementary Fig. 11b). To determine whether D-TLKIVW has affinity for soluble VQIVYK, we measured  $^1\text{H}$  NMR spectra of D-TLKIVW and D-LKTWIV in the presence of increasing amounts of soluble VQIVYK. Only a slight change in the respective chemical shifts of the indole proton peaks of D-TLKIVW and D-LKTWIV is observed, even at a 70-fold molar excess of VQIVYK (Supplementary Fig. 13). This, together with the ability of the peptide to prevent seeded fibril formation, suggests that D-TLKIVW does not interact with monomers but rather with a structured, fibril-like species.

As another test of our design model, we replaced the D-Leu residue with D-Ala in D-TLKIVW. Our structural model suggests that D-Leu 2 of D-TLKIVW is important for preventing tau fibril formation because of its favourable interaction with the Ile residue of the VQIVYK molecule below and with Ile and the first Val of VQIVYK across the steric zipper (Fig. 2b, c and Supplementary Fig. 1). The D-Ala replacement

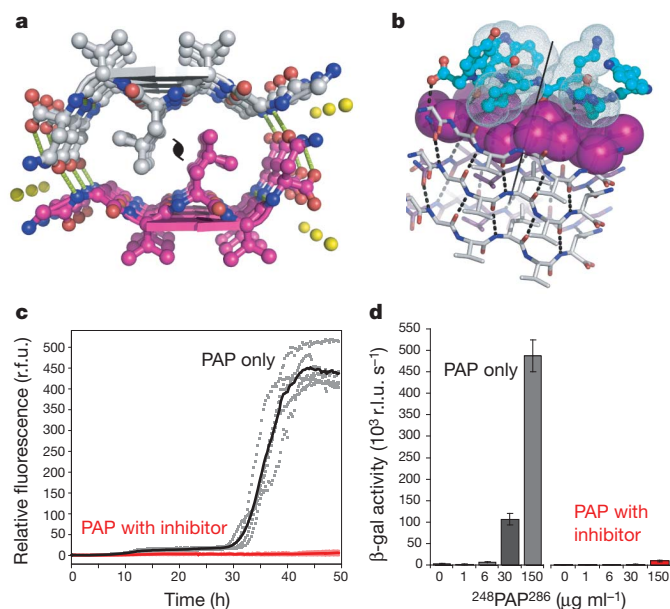
eliminates these interactions and, furthermore, removes a steric clash that would occur were another VQIVYK molecule placed across from the inhibitor (Supplementary Fig. 3 and Supplementary Table 1). When the D-Ala variant is incubated with VQIVYK and the tau constructs, it has no inhibitory effect on fibril formation (Fig. 2f and Supplementary Fig. 14). This confirms that D-Leu 2 is critical for the efficacy of D-TLKIVW, consistent with our model.

In summary, although our electron microscopy, NMR and D-Ala replacement results support a model in which the designed peptide D-TLKIVW binds to the ends of tau fibrils, they do not constitute proof that the inhibitors bind exactly as anticipated in the designs (Supplementary Fig. 15).

To expand on our design methodology, we computationally designed an inhibitor of  $^{248}\text{PAP}^{286}$  fibril formation containing non-natural L-amino acids (Fig. 4b and Supplementary Fig. 16), using the GGVLVN structure as a template (Fig. 4a and Supplementary Table 2). This peptide, Trp-His-Lys-chAla-Trp-hydroxyTic (WW61), contains an Ala derivative,  $\beta$ -cyclohexyl-L-alanine (chAla) and a Tyr/Pro derivative, 7-hydroxy-(S)-1,2,3,4-tetrahydroisoquinoline-3-carboxylic acid (hydroxyTic), both of which increase contact area with the GGVLVN template. The non-natural chAla forms hydrophobic interactions with the Leu residue in the steric-zipper interface, and hydroxyTic supports the favourable placement of chAla through hydrophobic packing (Fig. 4b and Supplementary Fig. 16b).



**Figure 3 | Mechanism of interaction.** **a**, Nanogold covalently bound to D-TLKIVW localizes at the ends (arrows) of two tau K19 fibrils. Scale bar, 50 nm. **b**, The inhibitor D-TLKIVW binds to fibrils with an estimated affinity constant in the low micromolar range, as shown by the indole proton region of the 500-MHz  $^1\text{H}$  NMR spectra of D-TLKIVW (9.83 p.p.m.) and D-LKTWIV (9.98 p.p.m.) in the presence of increasing concentrations of VQIVYK fibrils. The resonance of the D-TLKIVW indole proton is reduced in the presence of increasing concentrations of VQIVYK fibrils, whereas the indole proton signal for the scrambled control peptide D-LKTWIV is only slightly affected. Fibril solutions contained 0–1,500  $\mu\text{M}$  VQIVYK monomers, as indicated.



**Figure 4 | Designed non-natural peptide inhibits  $^{248}\text{PAP}^{286}$  fibril formation.** **a**, The view down the fibril axis of the crystal structure of the GGVLVN steric zipper reveals two mating  $\beta$ -sheets with parallel, in-register  $\beta$ -strands (hydrogen bonds, green dashed lines; water molecules, yellow spheres). **b**, View roughly perpendicular to a fibril of three layers, with the atoms of the side chains of the top layer shown as purple spheres. On top is a designed non-natural peptide inhibitor, Trp-His-Lys-chAla-Trp-hydroxyTic (blue; see Supplementary Fig. 16). **c**, The inhibitor blocks  $^{248}\text{PAP}^{286}$  fibril formation, as shown by monitoring Thioflavin T fluorescence. With a twofold molar excess of the inhibitor (pale red), the fluorescence remains low over the course of the experiment for all five replicates, unlike in the absence of inhibitor (grey). Mean fluorescence values are shown as solid red and black lines with and without the inhibitor, respectively. r.f.u., relative fluorescence units. **d**, HIV infection rates were determined by monitoring  $\beta$ -galactosidase ( $\beta$ -gal) activity. Agitated  $^{248}\text{PAP}^{286}$  alone efficiently increases viral infection, whereas  $^{248}\text{PAP}^{286}$  mixtures incubated with inhibitor were unable to enhance HIV infection. Peptide concentrations during virion treatment are indicated on the x axis. Error bars show the s.d. of three measurements per sample. r.l.u., relative light units.



Moreover, we propose that the bulky side chains and steric constraints of hydroxyTic provide hindrance to further fibril growth.

This designed peptide, WW61, effectively delays both seeded and unseeded fibril formation of  $^{248}\text{PAP}^{286}$  *in vitro* (Fig. 4c and Supplementary Figs 17 and 18). In the presence of a twofold molar excess of this inhibitor, seeded fibril formation is efficiently blocked for more than two days (Fig. 4c). Furthermore, we see that increasing the concentration of this inhibitor extends the fibril formation lag time (Supplementary Fig. 19). These inhibition assay results were further confirmed by electron microscopy (Supplementary Fig. 20). As a control for specificity, we tested the effect of GIHKQK, from the amino terminus of  $^{248}\text{PAP}^{286}$ , and PYKLWN, a peptide with the same charge as WW61. Neither peptide affected fibril formation kinetics, indicating that the inhibitory activity of the designed peptide is sequence specific (Supplementary Fig. 21).

Because  $^{248}\text{PAP}^{286}$  fibrils (SEVI) have been shown to enhance HIV infection<sup>22</sup>, using a functional assay we investigated whether WW61 is able to prevent this enhancement. In this experiment, we treated HIV particles with  $^{248}\text{PAP}^{286}$  solutions that had been agitated for 20 hours (to allow fibril formation) in the presence or absence of WW61, and infected TZM-bl indicator cells. As has been previously observed, SEVI efficiently enhanced HIV infection<sup>22</sup>. However,  $^{248}\text{PAP}^{286}$  incubated with the designed inhibitor prevented HIV infection (Fig. 4d).

We performed several control experiments to verify that the lack of infectivity observed in the assay is indeed due to the inhibition of SEVI formation. First we confirmed that in the absence of SEVI the designed inhibitor WW61 does not affect HIV infectivity (Supplementary Fig. 22a). We also found that the control peptides GIHKQK and PYKLWN, which do not inhibit  $^{248}\text{PAP}^{286}$  fibril formation, fail to decrease HIV infectivity (Supplementary Fig. 22b). Additionally, we observed that WW61 has no inhibitory effect on polylysine-mediated HIV infectivity<sup>27</sup>, further ruling out a non-specific electrostatic interaction mechanism (Supplementary Fig. 22a). Together, these results demonstrate that a peptide capable of preventing  $^{248}\text{PAP}^{286}$  fibril formation also inhibits the generation of virus-enhancing material.

Structure-based design of inhibitors of amyloid fibril formation has been challenging in the absence of detailed information about the atomic-level interactions that form the fibril spine. So far, one of the most successful structure-based approaches to preventing fibril formation has been to stabilize the native tetrameric structure of transthyretin<sup>28</sup>. That approach is well suited to the prevention of fibril formation of proteins with known native structures, but other proteins involved in amyloid-related diseases, such as tau protein, amyloid- $\beta$  and  $^{248}\text{PAP}^{286}$ , lack fully ordered native structures<sup>29</sup>. Our structure-based approach makes it possible to design inhibitors independent of native structure. Instead, the templates are atomic-level structures of short, fibril-forming segments<sup>14,15</sup>. By using these fibril-like templates, and adopting computational methods successful in designing novel proteins and protein-protein interfaces<sup>25,30</sup>, we have created specific inhibitors of proteins that normally form fibrils. These results support the hypothesis that the steric zipper is a principal feature of tau-related and SEVI fibrils, and suggest that, with current computational methods and steric-zipper structures, we have the tools to design specific inhibitors to prevent the formation of other amyloid fibrils.

## METHODS SUMMARY

We used crystal structures of hexapeptide segments of VQIVYK and GGVLVN as templates to design peptide inhibitors using the Rosetta software<sup>25</sup>. Briefly, this algorithm searches possible side-chain conformations (called rotamers) of all amino acids in a peptide  $\beta$ -strand backbone stacked onto the fibril end of both segment structures. The Rosetta software is extended to sample the approximate side-chain conformation of non-natural D and L amino acids by adapting side-chain torsion angles from those in their natural counterparts. The lowest energy set of side-chain rotamers is identified by combinatorial optimization of a potential consisting of a term for the Lennard-Jones potential, an orientation-dependent hydrogen-bond potential term, an implicit solvation term and a structure-derived side-chain and backbone torsional potential term.

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Received 6 December 2010; accepted 21 April 2011.**

**Published online 15 June 2011.**

1. Westermark, P. *et al.* A primer of amyloid nomenclature. *Amyloid* **14**, 179–183 (2007).
2. Maji, S. K. *et al.* Functional amyloids as natural storage of peptide hormones in pituitary secretory granules. *Science* **325**, 328–332 (2009).
3. Fowler, D. M., Koulou, A. V., Balch, W. E. & Kelly, J. W. Functional amyloid – from bacteria to humans. *Trends Biochem. Sci.* **32**, 217–224 (2007).
4. Astbury, W. T. & Dickinson, S. The X-ray interpretation of denaturation and the structure of the seed globulins. *Biochem. J.* **29**, 2351–2360 (1935).
5. Calamai, M., Chiti, F. & Dobson, C. M. Amyloid fibril formation can proceed from different conformations of a partially unfolded protein. *Biophys. J.* **89**, 4201–4210 (2005).
6. Tjernberg, L. O. *et al.* Arrest of  $\beta$ -amyloid fibril formation by a pentapeptide ligand. *J. Biol. Chem.* **271**, 8545–8548 (1996).
7. Findeis, M. A. Peptide inhibitors of  $\beta$  amyloid aggregation. *Curr. Top. Med. Chem.* **2**, 417–423 (2002).
8. Sciarretta, K. L., Gordon, D. J. & Meredith, S. C. Peptide-based inhibitors of amyloid assembly. *Methods Enzymol.* **413**, 273–312 (2006).
9. Soto, C., Kindy, M. S., Baumann, M. & Frangione, B. Inhibition of Alzheimer's amyloidosis by peptides that prevent  $\beta$ -sheet conformation. *Biochem. Biophys. Res. Commun.* **226**, 672–680 (1996).
10. Kokkoni, N., Stott, K., Amijee, H., Mason, J. M. & Doig, A. J. N-methylated peptide inhibitors of amyloid aggregation and toxicity. Optimization of the inhibitor structure. *Biochemistry* **45**, 9906–9918 (2006).
11. Sato, T. *et al.* Inhibitors of amyloid toxicity based on  $\beta$ -sheet packing of A $\beta$ 40 and A $\beta$ 42. *Biochemistry* **45**, 5503–5516 (2006).
12. Larbig, G., Pickhardt, M., Lloyd, D. G., Schmidt, B. & Mandelkow, E. Screening for inhibitors of tau protein aggregation into Alzheimer paired helical filaments: a ligand based approach results in successful scaffold hopping. *Curr. Alzheimer Res.* **4**, 315–323 (2007).
13. Wiesehan, K. *et al.* Selection of D-amino-acid peptides that bind to Alzheimer's disease amyloid peptide A $\beta$ <sub>1–42</sub> by mirror image phage display. *ChemBioChem* **4**, 748–753 (2003).
14. Nelson, R. *et al.* Structure of the cross- $\beta$  spine of amyloid-like fibrils. *Nature* **435**, 773–778 (2005).
15. Sawaya, M. R. *et al.* Atomic structures of amyloid cross- $\beta$  spines reveal varied steric zippers. *Nature* **447**, 453–457 (2007).
16. Wiltzius, J. J. *et al.* Molecular mechanisms for protein-encoded inheritance. *Nature Struct. Mol. Biol.* **16**, 973–978 (2009).
17. Selkoe, D. J. Alzheimer's disease: genes, proteins, and therapy. *Physiol. Rev.* **81**, 741–766 (2001).
18. Goux, W. J. *et al.* The formation of straight and twisted filaments from short tau peptides. *J. Biol. Chem.* **279**, 26868–26875 (2004).
19. von Bergen, M. *et al.* Assembly of  $\tau$  protein into Alzheimer paired helical filaments depends on a local sequence motif (<sup>306</sup>VQIVYK<sup>311</sup>) forming  $\beta$  structure. *Proc. Natl Acad. Sci. USA* **97**, 5129–5134 (2000).
20. Goldschmidt, L., Teng, P. K., Riek, R. & Eisenberg, D. Identifying the amyloids, proteins capable of forming amyloid-like fibrils. *Proc. Natl Acad. Sci. USA* **107**, 3487–3492 (2010).
21. Thompson, M. J. *et al.* The 3D profile method for identifying fibril-forming segments of proteins. *Proc. Natl Acad. Sci. USA* **103**, 4074–4078 (2006).
22. Münch, J. *et al.* Semen-derived amyloid fibrils drastically enhance HIV infection. *Cell* **131**, 1059–1071 (2007).
23. Friedhoff, P., von Bergen, M., Mandelkow, E. M., Davies, P. & Mandelkow, E. A nucleated assembly mechanism of Alzheimer paired helical filaments. *Proc. Natl Acad. Sci. USA* **95**, 15712–15717 (1998).
24. Wille, H., Drewes, G., Biernat, J., Mandelkow, E. M. & Mandelkow, E. Alzheimer-like paired helical filaments and antiparallel dimers formed from microtubule-associated protein tau *in vitro*. *J. Cell Biol.* **118**, 573–584 (1992).
25. Kuhlman, B. *et al.* Design of a novel globular protein fold with atomic-level accuracy. *Science* **302**, 1364–1368 (2003).
26. Chen, Z., Krause, G. & Reif, B. Structure and orientation of peptide inhibitors bound to  $\beta$ -amyloid fibrils. *J. Mol. Biol.* **354**, 760–776 (2005).
27. Roan, N. R. *et al.* The cationic properties of SEVI underlie its ability to enhance human immunodeficiency virus infection. *J. Virol.* **83**, 73–80 (2009).
28. Petrassi, H. M., Klabunde, T., Sacchetti, J. & Kelly, J. W. Structure-based design of N-phenyl phenoxazine transthyretin amyloid fibril inhibitors. *J. Am. Chem. Soc.* **122**, 2178–2192 (2000).
29. Schweers, O., Schonbrunn-Hanebeck, E., Marx, A. & Mandelkow, E. Structural studies of tau protein and Alzheimer paired helical filaments show no evidence for  $\beta$ -structure. *J. Biol. Chem.* **269**, 24290–24297 (1994).
30. Fleishman, S. J. *et al.* Computational design of proteins targeting the conserved stem region of influenza hemagglutinin. *Science* **332**, 816–821 (2011).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank M. I. Ivanova, J. Corn, T. Kortemme, D. Anderson, M. R. Sawaya, M. Phillips, S. Sambashivan, J. Park, M. Landau, A. Laganowsky, Q. Zhang, R. Clubb, F. Guo, T. Yeates, J. Nowick, J. Zheng and M. J. Thompson for discussions; the HHMI, NIH, NSF, Gates Foundation and Joint Center for Translational Medicine for support; R. Peterson for help with NMR experiments; E. Mandelkow for providing tau constructs; R. Riek for providing amyloid- $\beta$ ; and J. Stroud for amyloid- $\beta$  preparation.

Support came from the Damon Runyon Cancer Research Foundation (J.K.), the Ruth L. Kirschstein National Research Service Award (H.W.C.), the programme for junior professors by the Ministry of Science, Baden-Württemberg (J.M.), and a UCLA-IGERT bioinformatics traineeship (S.A.S.).

**Author Contributions** S.A.S., J.K., D.B., J.M. and D.E. designed the project. J.K. and S.A.S. created the design protocol. J.K. designed the  $\alpha$ -peptides. L.J. expanded the design methodology and designed the non-natural amino-acid peptides. S.A.S., H.W.C. and A.Z. performed the fluorescence experiments and electron microscopy, and analysed kinetic data. A.Z. determined the structure of GGVLVN. O.Z. performed the HIV

infectivity experiments. J.T.S. determined the tau fibril elongation rates. S.A.S. performed the NMR experiments. S.A.S., J.K. and D.E. wrote the manuscript and coordinated contributions by other authors.

**Author Information** Atomic coordinates and structure factors for the reported GGVLVN structure have been deposited in the Protein Data Bank with accession code 3PPD. Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at [www.nature.com/nature](http://www.nature.com/nature). Correspondence and requests for materials should be addressed to D.E. ([david@mbi.ucla.edu](mailto:david@mbi.ucla.edu)).

## METHODS

**Computational design.** Computational designs were carried out using the Rosetta software<sup>25</sup> (<http://www.rosettacommons.org>). This algorithm involves building side-chain rotamers of all amino acids onto a fixed protein backbone. The lowest energy set of side-chain rotamers is then identified as those which minimize an energy function containing a Lennard–Jones potential, an orientation-dependent hydrogen-bond potential, a solvation term, amino-acid-dependent reference energies and a statistical torsional potential that depends on the backbone and side-chain dihedral angles.

**D-amino-acid tau inhibitors.** The crystal structure of VQIVYK (ref. 15; Protein Data Bank ID, 2ON9) was used as a starting scaffold for computational design. To take full advantage of the statistical nature of the rotamer library and some terms in the Rosetta energy function, the stereochemistry of the fibril scaffold was inverted so that design would take place using L amino acids. An extended L-peptide was aligned with the N, C and O backbone atoms of the D-fibril scaffold. This L-peptide was subsequently redesigned, keeping all atoms of the D-fibril fixed. The stereochemistry of the final design model was then inverted, yielding a D-peptide designed to cap an L-fibril. We inspected the finished models to confirm that inversion of the stereochemistry at the Thr and Ile C $\beta$  atoms did not make the designs energetically unfavourable. Energetic consequences of incorporating a D inhibitor peptide in the middle of an L fibril were subsequently evaluated to ensure that fibril propagation could not continue after association of an inhibitor. Calculations of the area buried and shape complementarity were performed with AREAIMOL<sup>31</sup> and SC<sup>32</sup>, respectively.

**L-peptide <sup>248</sup>PAP<sup>286</sup> inhibitors.** The crystal structure of GGVLVN (PDB ID, 3PPD) was used as a template for the following design procedure. An extended L-peptide was aligned according to crystal symmetry. Small, random perturbations of the L-peptide were performed to optimize the rigid-body arrangement between the fibril template and the peptide inhibitor. Full sequence optimization of the inhibitor was performed using the Rosetta software package, allowing residues directly contacting the inhibitor to repack; other scaffold residues remained fixed. Because the design calculations use a discrete rotamer representation of the side chains, we next performed simultaneous quasi-Newtonian optimization of the inhibitor rigid-body orientation, the side-chain torsion angles and, in some cases, the backbone torsion angles using the full-atom Rosetta energy function. This optimization was essential to the subsequent assessment of the inhibition of the design. Several iterative runs of small perturbations in inhibitor placement, interface design and refinement were performed to improve hydrogen-bonding and packing interactions. The designs that ranked highest on the basis of the total binding energy between the inhibitor and the fibril scaffold and the interfacial shape complementarity<sup>32</sup> were subsequently synthesized and tested.

For each initial active L-peptide design, the non-natural L amino acids were incorporated using a growth strategy. Non-natural amino acids, structurally similar to those of initial active designs, were selected on the basis of their solubility, side-chain shape and commercial availability. Side-chain conformations were approximately sampled by adopting side-chain torsion angles from those in their natural counterparts. Sequence optimization of the inhibitor was performed and the optimal set of rotamers identified using Monte Carlo simulated annealing with the full-atom energy function described above. The resulting designs were ranked on the basis of the total binding energy between the inhibitor and the fibril scaffold.

**Tau construct expression and purification.** pNG2 expression vectors (derived from pET-3b<sup>33</sup>) containing either the K12 or K19 gene were provided by E. Mandelkow<sup>34</sup>. Expression in BL21(DE3) *Escherichia coli*<sup>33</sup> was induced with 1 mM isopropyl thiogalactoside when the absorbance  $A_{600\text{ nm}}$  was between 0.8 and 1.0, and cells were collected after 3–4 h. K12 and K19 were purified on the basis of previously described methods<sup>35</sup>. Cells were pelleted for 20 min at 4,700g and resuspended in 20 mM MES, pH 6.8, 1 mM EDTA, 0.2 mM MgCl<sub>2</sub>, 5 mM DTT, 1 mM PMSF and a protease inhibitor cocktail. The cells were sonicated for 2.5 min and, following addition of NaCl to bring cell lysate to 0.5 M NaCl, the lysate was boiled for 20 min. The lysate was sedimented at 30,000g for 20 min and dialyzed twice against 20 mM MES, pH 6.8, 50 mM NaCl, 1 mM EDTA, 1 mM MgCl<sub>2</sub>, 2 mM DTT and 0.1 mM PMSF at 4 °C. The dialysate was pelleted for 20 min at 30,000g and filtered before cation exchange chromatography on an AKTA Explorer (GE Pharmacia) with a HighTrap HP SP 5-ml column (GE Healthcare). The sample was eluted with a linear gradient of up to 60% buffer B (20 mM MES, pH 6.8, 1 M NaCl, 1 mM EDTA, 1 mM MgCl<sub>2</sub>, 2 mM DTT and 0.1 mM PMSF). Size exclusion chromatography was optionally performed with a Superdex 75 10/300 GL column (GE Healthcare) in PBS buffer (137 mM NaCl, 3 mM KCl, 10 mM Na<sub>2</sub>HPO<sub>4</sub>, 2 mM KH<sub>2</sub>PO<sub>4</sub>, pH 7.4) with 1 mM DTT on the AKTA Explorer depending on preparation purity as assessed by SDS polyacrylamide gel electrophoresis.

**Tau construct inhibition assays.** Fibril formation assays were performed on the basis of previously published protocols<sup>35–38</sup>. Reaction mixtures (150  $\mu$ l) containing

50  $\mu$ M tau K12 or K19, as determined by the Micro BCA Protein Assay Kit (Pierce), were incubated in 250 mM sodium phosphate buffer, pH 7.4, with 1 mM DTT, 12.5  $\mu$ M heparin (average molecular mass, 6,000 Da; Sigma) and 10  $\mu$ M Thioflavin S (ThS; MP Bio). Inhibitor peptides (CS Bio, Celtek Biosciences) were dissolved in 250 mM phosphate buffer, pH 7.4, to 0.5 mM and added at specified molar ratios. Reactions were split into a minimum of three replicates in black, 96-well, optically clear plates (Nunc), sealed with Corning pressure-sensitive sealing tape and monitored using either a Varioskan plate reader (Thermo Scientific), for K12, or a SpectraMax M5, for K19. The fluorescence signal was measured every 15 min with excitation and emission wavelengths of 440 and 510 nm, respectively, at 37 °C, with continuous shaking at 900 r.p.m. with a diameter of 1 mm for K12, and with quiescent incubation with shaking 2 s before each reading for K19. Plots showing the fluorescence trace of the replicate with median lag time for each sample were created using R<sup>39</sup>. Plots of lag time depict the mean time value at which each replicate crossed an arbitrary fluorescence value above noise background (values were selected per experiment and applied to all samples). Error bars represent the standard deviation of the replicate lag times for each sample.

**Seeded K12 fibril formation assays.** Seeds were produced by incubating 50  $\mu$ M K12 as above, but without ThS present, and were added at 0.25% (v/v). Peptide stock concentrations were 0.75 mM and were added at a final concentration of tenfold molar excess relative to soluble K12. Reaction mixtures were otherwise prepared and monitored as above.

**VQIVYK inhibition assays.** The VQIVYK fibril formation assay was modified from a previously published protocol<sup>40</sup>. Buffers and plates were kept on ice to delay VQIVYK fibril formation while the reaction mixtures were prepared. Replicate solutions of 180  $\mu$ l of 25 mM MOPS, pH 7.2, 100  $\mu$ M ThS and inhibitor peptides were added to black, clear-bottomed, 96-well Nunc plates with 1/8-inch PTFE beads (Orange Products). Acetylated and amidated VQIVYK (Genscript) was dissolved in H<sub>2</sub>O to 1.3 mM and filtered through a Millipore Microcon 100-kDa filter device at 14,000g for 5 min at 4 °C to remove large aggregates (final concentration, ~1 mM). Filtered VQIVYK (20  $\mu$ l) was added to each reaction well. ThS fluorescence was monitored at room temperature every 2 min using a SpectraMax M5 fluorometer with 2 s of mixing before each reading.

**Amyloid- $\beta$  fibril formation assay.** Lyophilized amyloid- $\beta$ (1–42) was diluted to 0.2 mg ml<sup>–1</sup> in 50 mM NH<sub>4</sub>OH and filtered with a 0.2- $\mu$ m filter. The reaction mixture contained a final concentration of 11.5  $\mu$ M amyloid- $\beta$ (1–42), 10  $\mu$ M Thioflavin T (ThT), 23 mM NH<sub>4</sub>OH in 100 mM bicine, pH 9.1, and 11.5  $\mu$ M D-TLKIVW in reactions with peptide present. Reactions were split into four replicates and the ThT fluorescence signal was measured every minute (excitation wavelength, 440 nm; emission wavelength, 510 nm), at 37 °C, with continuous shaking at 960 r.p.m. with a 1-mm diameter in a Varioskan fluorometer.

**Electron microscopy.** Sample (5  $\mu$ l) was applied to glow-discharged, 400-mesh carbon-coated, formvar films on copper grids (Ted Pella) for 3 min. Grids were rinsed twice with distilled water and stained with 1% uranyl acetate for 90 s. Grids were examined in a Hitachi H-7000 transmission electron microscope at 75 keV or a JEOL JEM1200-EX operating at 80 keV.

**Tau fibril formation kinetic analysis.** The nucleation ( $k_1$ ) and propagation ( $k_2$ ) rates were determined by fitting the form of the Finke–Watzky two-step mechanism<sup>41</sup>. Plateau values were determined and the remaining parameters were fitted using the ‘leasqr’ nonlinear least-squares regression function (<http://fly.isti.cnr.it/pub/software/octave/leasqr/>) through the OCTAVE software package (<http://www.gnu.org/software/octave/>).

**Preparation of peptide–gold conjugates.** Peptide–Nanogold conjugates were prepared as described earlier for similarly sized peptides<sup>42</sup>. Briefly, 60 nmol of the peptides CGGG-(D)-TLKIVW and CGGG-(D)-LKTWIV (CS Bio) were dissolved in 110  $\mu$ l of phosphate-buffered saline (20 mM, pH 6.5, 0.15 M NaCl), added to 6 nmol of Monomaleimido Nanogold (Nanoprobe), dissolved in 200  $\mu$ l H<sub>2</sub>O and incubated for 1 h at room temperature (22 °C) with constant rotation. Peptide–Nanogold conjugates were separated from excess unbound peptides by membrane centrifugation (Microcon-10 system, Amicon) using a molecular mass cut-off of 10 kDa. Peptide–Nanogold conjugates were then diluted into phosphate-buffered saline, aliquoted and stored at –20 °C for no longer than one month.

**Preparation of K19 fibrils.** K19 fibrils were generated by incubating 100  $\mu$ M soluble K19 with 25  $\mu$ M 6-kDa heparin overnight at 37 °C in phosphate buffer (50 mM, pH 7.4). K19 fibrils were sonicated for 15 s, using a microtip set to 35% amplitude. Residual heparin and small oligomers were removed by centrifuging the mixture through a 100-kDa Microcon concentrator for 10 min at 14,000g, washing the retentate with phosphate buffer and repeating three times; the retentate was restored to its original volume with phosphate buffer. These short fibril segments were stored at 4 °C for no longer than one week. For NMR studies, fibril samples were similarly prepared, but were washed in H<sub>2</sub>O and concentrated to 2 mM K19 (by monomer).



**Preparation of samples for Nanogold binding experiments.** Nanogold conjugated inhibitor (or control) (10 nM) was incubated with 1.67  $\mu\text{M}$  K19 fibrils (by monomer) in MOPS buffer (25 mM, pH 7.2) for 1 h. We applied 5  $\mu\text{l}$  of it to a glow-discharged, 400-mesh carbon-stabilized copper grid (Ted Pella) for 3 min. The grids were washed twice with  $\text{H}_2\text{O}$  and 10  $\mu\text{l}$  of the Goldenhance reagent was applied for 10 s. The grids were washed five times with  $\text{H}_2\text{O}$  and negatively stained with 2% uranyl acetate.

**Quantification and localization of Nanogold binding.** For each sample, 75 Nanogold particles  $\leq 15$  nm in diameter were counted and classified as bound or unbound. The 15-nm cut-off was chosen to exclude unbound, but adjacent, particles enlarged by Goldenhance that only apparently bind fibrils. To establish the localization of the binding observed, individual Nanogold particles bound to fibrils were categorized as bound to the fibril end or side. In both of these experiments, sample identities were concealed from the microscopist to ensure unbiased counting. Grids were examined with a JEOL JEM1200-EX and images were recorded using DIGITALMICROGRAPH (Gatan).

**Statistical analysis of Nanogold binding.** We compared counts of Nanogold-conjugated peptides and unconjugated Nanogold bound to fibrils or localizing to fibril ends. Twenty-one unconjugated Nanogold particles out of 75 counted bound to fibrils. We modelled Nanogold particles bound to fibrils using a binomial distribution with parameters  $n = 75$  (sample size: number of observations) and  $P = 0.28$  (probability of success). In a separate experiment, 22 unconjugated Nanogold particles bound to fibrils that localized to fibril ends, following a binomial distribution with  $n = 105$  and  $P = 0.21$ .

Because the number of counts is fairly large, we assumed a normal distribution and used a standard Z-test to compare the number of bound Nanogold-peptide conjugates with the expected distribution based on the number of bound, unconjugated Nanogold particles. We used an analogous analysis to determine the significance of localization to fibril ends.

The numbers of Nanogold-D-TLKIVW conjugates bound to fibrils ( $x_{\text{bound}} = 43$ ,  $n = 75$ ) and bound Nanogold-D-TLKIVW conjugates localizing to the end of fibrils ( $x_{\text{end}} = 49$ ,  $n = 86$ ) were significantly different from the corresponding numbers for Nanogold alone, whereas the number of Nanogold-D-LKTWIV conjugates bound ( $x_{\text{bound}} = 15$ ,  $n = 75$ ) or the number localized to fibril ends ( $x_{\text{end}} = 17$ ,  $n = 100$ ) did not differ significantly from the corresponding numbers for Nanogold alone.

**VQIVYK preparation for binding studies.** Acetylated and amidated VQIVYK peptide (Genscript) was dissolved to 1 mM in 25 mM MOPS, pH 7.2, and incubated at room temperature for at least 24 h. Fibrils were washed with  $\text{H}_2\text{O}$ , concentrated using an Amicon ultracentrifugal filter with a 3-kDa molecular mass cut-off and resuspended in  $\text{H}_2\text{O}$  to a final concentration (by monomer) of 4 mM. Soluble VQIVYK was prepared by dissolving VQIVYK peptide (CS Bio) with free amino and carboxy termini in  $\text{H}_2\text{O}$ .

**$^1\text{H}$  NMR sample preparation and measurements.** NMR samples were prepared with 5%  $\text{D}_2\text{O}$  and 10 mM NaOAc, pH 5.0. D-peptides were added from 1 mM stocks in  $\text{H}_2\text{O}$  to a final concentration of 100  $\mu\text{M}$ . Soluble and fibrillar VQIVYK and tau protein were added at indicated concentrations to make a final volume of 550  $\mu\text{l}$ .  $^1\text{H}$  NMR spectra measured at 500 MHz were collected on a Bruker DRX500 at 283 K.  $\text{H}_2\text{O}$  resonance was suppressed through presaturation. Spectra were processed with XWINNMR 3.6.

**Binding constant estimations.** NMR data were analysed to estimate a binding constant for the interaction between D-TLKIVW and VQIVYK fibrils. At about 1,000  $\mu\text{M}$  VQIVYK (concentration as monomer), 50% of D-TLKIVW is bound (Supplementary Fig. 11). The steric-zipper model suggests that there are two monomers per 4.7 Å (0.47-nm) layer in a fibril<sup>14</sup>—such that the number of monomers per fibril is given by [fibril length (nm)]  $\times$  (2 monomers per 0.47 nm)—and we estimate the fibril concentration using the monomer concentration:  $[\text{VQIVYK}_{\text{fibril}}] = [\text{VQIVYK}_{\text{monomer}}]/(\text{monomers per fibril})$ . If we assume one binding site and estimate from electron microscopy an average length of  $\sim 140$  nm per fibril, then there are about 600 monomers per fibril, and the apparent dissociation constant is about 2  $\mu\text{M}$ .

**Hydrogen-bonding energy calculation.** We used AREAIMOL<sup>31</sup> to calculate the non-polar and polar areas buried by the interaction between D-TLKIVW with the VQIVYK steric zipper (Fig. 2b, c and Supplementary Fig. 1). We calculate buried areas of 201, 24 and 102 Å<sup>2</sup> for carbon, nitrogen and oxygen atoms, respectively. Using the atomic solvation parameters of ref. 43, we estimate that the free energy of transferring the inhibitor from a non-polar phase to an aqueous phase,  $\Delta G_{\text{solvation}}$ , is approximately 2.5 kcal mol<sup>-1</sup>. On the basis of an apparent dissociation constant of 2  $\mu\text{M}$ , we estimate the total free energy change of bringing the inhibitor into contact with the VQIVYK steric-zipper template,  $\Delta G_{\text{binding}}$ , to be 7.4 kcal mol<sup>-1</sup>. From the interaction model (Fig. 2c and Supplementary Fig. 1), we maintain six hydrogen bonds between D-TLKIVW and VQIVYK, and estimate the free energy change per hydrogen bond to be  $(\Delta G_{\text{binding}} - \Delta G_{\text{solvation}})/6$ , or  $\sim 0.8$  kcal mol<sup>-1</sup>.

**GGVLVN crystallization and structure determination.** The GGVLVN peptide was dissolved in 10 mM Tris, pH 9, at 1.8 mg ml<sup>-1</sup> and crystallized in 10% (w/v) PEG-8000, 0.1 M MES, pH 6.0, and 0.2 M Zn(OAc)<sub>2</sub>. X-ray diffraction data was collected at APS beamline 24-ID-E. Phases were determined by molecular replacement using an idealized  $\beta$ -strand in PHASER<sup>44</sup>. Crystallographic refinement was performed using REFMAC<sup>45</sup>. Model building was performed with COOT<sup>46</sup> and illustrated with PYMOL<sup>47</sup>.

**<sup>248</sup>PAP<sup>286</sup> fibril formation and inhibition.** Fmoc- $\beta$ -cyclohexyl-L-alanine and Fmoc-7-hydroxy-(S)-1,2,3,4-tetrahydroisoquinoline-3-carboxylic acid were purchased from AnaSpec and the inhibitor peptide Trp-His-Lys-chAla-Trp-hydroxyTic (WW61) was synthesized by Celtek Biosciences. <sup>248</sup>PAP<sup>286</sup> and WW61 were dissolved as  $\times 1.25$  and  $\times 5$  stocks in PBS, respectively, and filtered with a 0.1- $\mu\text{m}$  filter. <sup>248</sup>PAP<sup>286</sup> was diluted with PBS to 0.66 mM and ThT was added to 10  $\mu\text{M}$  final concentration. Samples were optionally mixed with 1.32 mM WW61 and vortexed. Five replicates of 150  $\mu\text{l}$  were immediately dispensed into a 96-well plate. In dose-response experiments, WW61 final concentrations were 0.33, 0.66 and 1.32 mM. Plates were continuously agitated at 960 r.p.m. at 37 °C, and ThT fluorescence readings were recorded (excitation wavelength, 440 nm; emission wavelength, 482 nm) at 15-min intervals with a Varioskan Flash fluorometer. Lag time was determined when fluorescence crossed an arbitrary value (3 r.f.u.) above background.

**Effect of WW61 on fibril-mediated enhancement of HIV-1 infection.** The CCR5 tropic molecular HIV-1 clone NL4\_3/92TH014-2<sup>48</sup> was generated by transient transfection of 293T cells with proviral DNA. Supernatants were collected 48 h later and p24 concentrations determined by ELISA. TZM-bl reporter cells encoding a lacZ gene under the control of the viral LTR promoter were obtained through the NIH AIDS Research and Reference Reagent Program and provided by Dr John C. Kappes, Dr Xiaoyun Wu and Tranzyme<sup>49</sup>. HIV-1 (40  $\mu\text{l}$ ) containing 0.1 ng of p24 antigen was incubated with 40- $\mu\text{l}$  dilutions of mixtures of <sup>248</sup>PAP<sup>286</sup> and inhibitory peptide, WW61, that was either freshly prepared or had been agitated for 23 h. Peptide concentrations and experimental conditions during agitation were similar to those described above. Thereafter, 20  $\mu\text{l}$  of the mixtures were used to infect 180  $\mu\text{l}$  of TZM-bl cells seeded the day before ( $10^5$  per well). Two days later, infection rates were determined by quantifying  $\beta$ -galactosidase activities in cellular lysates using the Gal-Screen assay (Applied Biosystems, T1027). Luminescence was recorded on an Orion microplate luminometer as relative light units per second.

**Effect of WW61 on polylysine-mediated enhancement of HIV-1 infection.** Polylysine (Sigma Aldrich) (50  $\mu\text{l}$ ) was mixed with an equal volume of WW61. Thereafter, 35- $\mu\text{l}$  fivefold dilutions of the polylysine-WW61 mixture or polylysine alone were incubated with the same volume of virus and incubated for 5 min at room temperature. Polylysine-WW61 concentrations were 100, 20, 4, 0.8, 0.16, 0.032, 0.064 and 0  $\mu\text{g ml}^{-1}$  during pre-incubation with virus stocks. Thereafter, 20  $\mu\text{l}$  of each mixture was added to 180  $\mu\text{l}$  of TZM-bl cells. The infection rate was determined two days later as described above.

**Effect of WW61, GIHKQK and PYKLWN on HIV-1 infection.** Each peptide (40  $\mu\text{l}$ ) was incubated with an equal volume of virus containing 1 ng of p24 antigens for 5 min at room temperature. Peptide concentrations were 150, 30, 6, 1.2 and 0  $\mu\text{g ml}^{-1}$  during pre-incubation with virus stocks. Thereafter, 20  $\mu\text{l}$  of each mixture was added separately to 180  $\mu\text{l}$  of TZM-bl cells (tenfold dilution) and the infection rate was determined as above.

- Collaborative Computational Project, Number 4. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D* **50**, 760–763 (1994).
- Lawrence, M. C. & Colman, P. M. Shape complementarity at protein/protein interfaces. *J. Mol. Biol.* **234**, 946–950 (1993).
- Studier, F. W., Rosenberg, A. H., Dunn, J. J. & Dubendorff, J. W. Use of T7 RNA polymerase to direct expression of cloned genes. *Methods Enzymol.* **185**, 60–89 (1990).
- Biernat, J. et al. The switch of tau protein to an Alzheimer-like state includes the phosphorylation of two serine-proline motifs upstream of the microtubule binding region. *EMBO J.* **11**, 1593–1597 (1992).
- Barghorn, S., Biernat, J. & Mandelkow, E. Purification of recombinant tau protein and preparation of Alzheimer-paired helical filaments in vitro. *Methods Mol. Biol.* **299**, 35–51 (2005).
- Friedhoff, P., Schneider, A., Mandelkow, E. M. & Mandelkow, E. Rapid assembly of Alzheimer-like paired helical filaments from microtubule-associated protein tau monitored by fluorescence in solution. *Biochemistry* **37**, 10223–10230 (1998).
- Pérez, M., Valpuesta, J. M., Medina, M., Montejó de Garcini, E. & Avila, J. Polymerization of tau into filaments in the presence of heparin: the minimal sequence required for tau-tau interaction. *J. Neurochem.* **67**, 1183–1190 (1996).
- Schweers, O., Mandelkow, E. M., Biernat, J. & Mandelkow, E. Oxidation of cysteine-322 in the repeat domain of microtubule-associated protein tau controls the in vitro assembly of paired helical filaments. *Proc. Natl Acad. Sci. USA* **92**, 8463–8467 (1995).

39. R Development Core Team. *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, Vienna) (<http://www.r-project.org>) (2008).
40. Rojas Quijano, F. A., Morrow, D., Wise, B. M., Brancia, F. L. & Goux, W. J. Prediction of nucleating sequences from amyloidogenic propensities of tau-related peptides. *Biochemistry* **45**, 4638–4652 (2006).
41. Morris, A. M., Watzky, M. A., Agar, J. N. & Finke, R. G. Fitting neurological protein aggregation kinetic data via a 2-step, minimal “Ockham’s razor” model: the Finke-Watzky mechanism of nucleation followed by autocatalytic surface growth. *Biochemistry* **47**, 2413–2427 (2008).
42. Schmidt, K., Segond von Banchet, G. & Heppelmann, B. Labelling of peptides with 1.4-nm gold particles to demonstrate their binding sites in the rat spinal cord. *J. Neurosci. Methods* **87**, 195–200 (1999).
43. Eisenberg, D., Wesson, M. & Yamashita, M. Interpretation of protein folding and binding with atomic solvation parameters. *Chem. Scr.* **29A**, 217–221 (1989).
44. McCoy, A. J. *et al.* Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674 (2007).
45. Murshudov, G. N., Vagin, A. A. & Dodson, E. J. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D* **53**, 240–255 (1997).
46. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **60**, 2126–2132 (2004).
47. DeLano, W. L. *PyMOL Molecular Viewer* (<http://www.pymol.org>) (2002).
48. Papkalla, A., Munch, J., Otto, C. & Kirchhoff, F. Nef enhances human immunodeficiency virus type 1 infectivity and replication independently of viral coreceptor tropism. *J. Virol.* **76**, 8455–8459 (2002).
49. Platt, E. J., Wehrly, K., Kuhmann, S. E., Chesebro, B. & Kabat, D. Effects of CCR5 and CD4 cell surface concentrations on infections by macrophagetropic isolates of human immunodeficiency virus type 1. *J. Virol.* **72**, 2855–2864 (1998).

# Structural insight into brassinosteroid perception by BRI1

Ji She<sup>1,2,3\*</sup>, Zhifu Han<sup>1\*</sup>, Tae-Wuk Kim<sup>4</sup>, Jinjing Wang<sup>3</sup>, Wei Cheng<sup>3</sup>, Junbiao Chang<sup>5</sup>, Shuai Shi<sup>5</sup>, Jiawei Wang<sup>1</sup>, Maojun Yang<sup>1</sup>, Zhi-Yong Wang<sup>4</sup> & Jijie Chai<sup>1,3</sup>

Brassinosteroids are essential phytohormones that have crucial roles in plant growth and development. Perception of brassinosteroids requires an active complex of BRASSINOSTEROID-INSENSITIVE 1 (BRI1) and BRI1-ASSOCIATED KINASE 1 (BAK1). Recognized by the extracellular leucine-rich repeat (LRR) domain of BRI1, brassinosteroids induce a phosphorylation-mediated cascade to regulate gene expression. Here we present the crystal structures of BRI1(LRR) in free and brassinolide-bound forms. BRI1(LRR) exists as a monomer in crystals and solution independent of brassinolide. It comprises a helical solenoid structure that accommodates a separate insertion domain at its concave surface. Sandwiched between them, brassinolide binds to a hydrophobicity-dominating surface groove on BRI1(LRR). Brassinolide recognition by BRI1(LRR) is through an induced-fit mechanism involving stabilization of two interdomain loops that creates a pronounced non-polar surface groove for the hormone binding. Together, our results define the molecular mechanisms by which BRI1 recognizes brassinosteroids and provide insight into brassinosteroid-induced BRI1 activation.

Ubiquitously distributed throughout the plant kingdom, brassinosteroids are a class of low-abundance phytohormones that have crucial roles in many aspects of growth and development<sup>1,2</sup>. Studies in *Arabidopsis* have led to the identification of a number of genes involved in brassinosteroid perception and signalling<sup>3–9</sup>. One of them, *BRI1*, when mutated, abolishes brassinosteroid-mediated responses of plants<sup>3</sup>. *BRI1* belongs to a large family of plant LRR receptor-like kinases (RLKs) with more than 200 members in *Arabidopsis* and over 300 in rice<sup>10</sup>. LRR RLKs are characterized by an extracellular LRR domain, a single-pass transmembrane segment and a cytoplasmic kinase domain. *BRI1* has been established as a bona fide receptor of brassinosteroids by genetic and biochemical investigations<sup>3,11,12</sup>. The most convincing evidence for this comes from a biochemical study showing that BRI1(LRR) is essential and sufficient for recognition of brassinosteroids<sup>12</sup>. A 70-residue island domain (residues 587–656) is indispensable for brassinosteroid recognition. Binding of brassinosteroids to BRI1 initiates a phosphorylation-mediated cascade, transducing the extracellular steroid signal to transcriptional programs<sup>13,14</sup>.

A second gene named *BAK1* involved in brassinosteroid signalling also encodes an LRR RLK, albeit with only five extracellular LRRs<sup>4,5</sup>. Although not involved in brassinosteroid binding, *BAK1* promotes brassinosteroid-induced signalling by physically associating with BRI1 (refs 4, 5, 15). Several models<sup>2,4,5,8,16–19</sup> of brassinosteroid-induced BRI1 activation converge at transphosphorylation within the BRI1–BAK1 complex, which involves brassinosteroid-enhanced BRI1 homodimerization<sup>16</sup>, release of BRI1 KINASE INHIBITOR 1 (BK1)<sup>19</sup> from BRI1, and BRI1–BAK1 heterodimerization<sup>17–19</sup>. It is still not fully understood how brassinosteroids are perceived by BRI1 to give rise to these events.

In the present study, we report the crystal structures of free BRI1(LRR) and its complex with brassinolide (the most active form of brassinosteroid) (Supplementary Table 1). The structures not only reveal the molecular mechanisms underlying brassinosteroid recognition by BRI1 and provide insight into brassinosteroid-induced BRI1

activation, but also explain the structure–activity relationship of brassinosteroids and serve as a foundation for the rational design of nonsteroidal mimetics.

## A helical solenoid structure of BRI1(LRR)

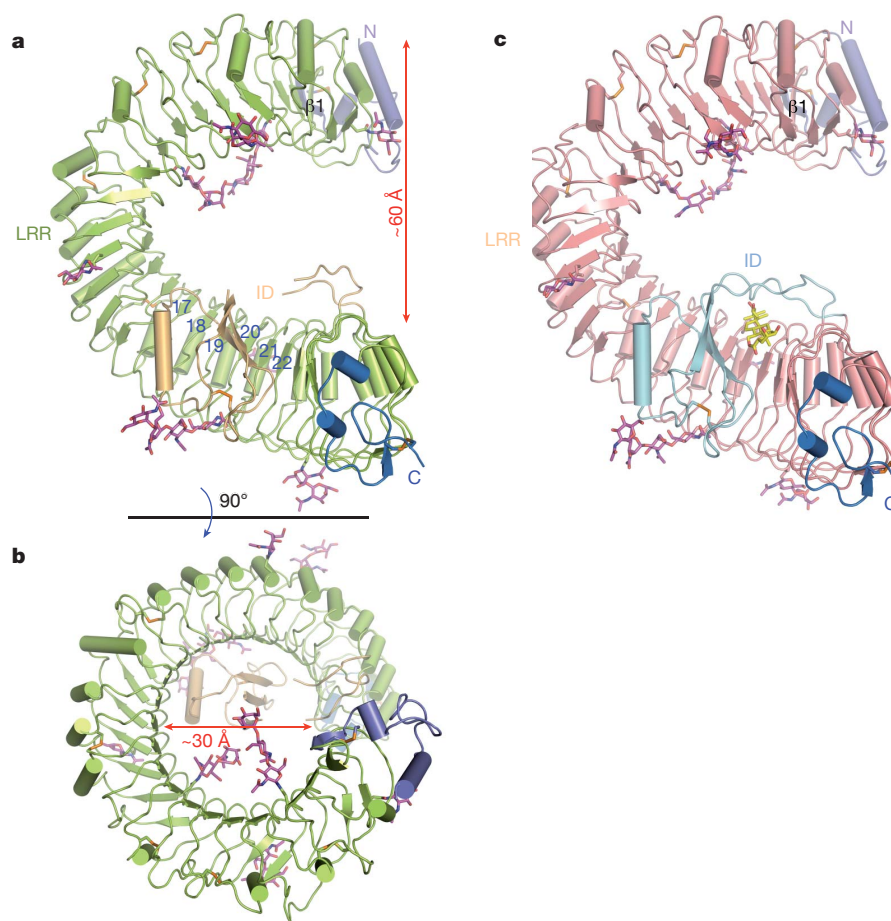
The three-dimensional structure of brassinolide-free BRI1(LRR) (residues 30–589, 596–643 and 647–772) contains 25 LRRs as predicted<sup>3</sup>, including 24 regular ones and an irregular one abutting the amino-terminal side. The LRRs packed in tandem assemble into a highly curved solenoid structure (Fig. 1a), with the overall rotation angle about the central axis approximately 360 degrees (Fig. 1b). Compared to the regular horseshoe-shaped structures of other LRR proteins, one notably distinct feature of the BRI1(LRR) structure is that it is exceptionally twisted, resulting in formation of a whole turn of right-handed helix with an inner diameter of about 30 Å. Relative to LRR1, LRR25 shifts about 60 Å along the central axis of the helix (Fig. 1a). An unprecedented structural feature of BRI1(LRR) is that it has an insertion domain anchored to the inner surface of the solenoid and running across six LRRs (Fig. 1a). The N-terminal capping domain (residues 30–70), consisting of one  $\beta$  strand ( $\beta$ 1) and two  $\alpha$  helices, is integrated into the LRR structure by forming an anti-parallel  $\beta$  sheet with the  $\beta$  strand from the irregular repeat, whereas the carboxy-terminal capping domain (residues 752–772) uses two short helices to tightly pack against the last repeat (Fig. 1a). Eight potential glycosylation sites (Supplementary Fig. 2) as defined by sufficient electron density were found in BRI1(LRR). An assignment of functions to them awaits further investigations.

Like canonical LRR proteins, the concave surface of the BRI1(LRR) solenoid is composed of a parallel  $\beta$  sheet comprising 25 continuously running parallel  $\beta$  strands. Compared to many other LRR proteins, however, BRI1(LRR) also possesses parallel but distorted  $\beta$  sheets following the inner  $\beta$  structure in most of the repeats (Fig. 1a, b). This is probably due to the plant-specific consensus sequence L/FGX<sub>1</sub>/vP (X and x stand for polar and any amino acid, respectively)<sup>20</sup> of BRI1(LRR)s (Supplementary Fig. 2) found in many LRR RLKs, as a similar structural

<sup>1</sup>Key Laboratory for Protein Sciences of Ministry of Education School of Life Sciences, Tsinghua University, Beijing 100084, China. <sup>2</sup>College of Biological Sciences, Peking University, Beijing 100084, China. <sup>3</sup>National Institute of Biological Sciences, No. 7 Science Park Road, Beijing 102206, China. <sup>4</sup>Department of Plant Biology, Carnegie Institution for Science, Stanford, California 94305, USA. <sup>5</sup>Department of Chemistry, Zhengzhou University, Zhengzhou 450001, China.

\*These authors contributed equally to this work.





**Figure 1 | BRI1(LRR) has a helical solenoid structure.** **a, b**, Overall structures of brassinolide-free BRI1(LRR) shown in two different orientations. The N-linked sugars (N-acetylglucosamines) are shown in magenta stick representation. Coloured in orange are disulphide bonds. ID, insertion domain. The N- and C-terminal cap is shown in slate grey and marine blue, respectively.

feature also exists in the plant LRR protein PGIP2 (ref. 21). In contrast with the concave side, the convex outer surface consists of varied structure elements, including  $3_{10}$  helices,  $\alpha$  helices and different length of loops (Fig. 1a, b). Some of the loops are stabilized through disulphide bonds formed between two consecutive repeats (LRR2–LRR3, LRR5–LRR6, LRR7–LRR8, LRR10–LRR11 and LRR14–LRR15) (Fig. 1a and Supplementary Fig. 1). The overall structure of BRI1(LRR) (residues 30–775) remains nearly unchanged on brassinolide binding, with a root mean squared deviation (r.m.s.d.) of 0.70 Å over 734 C $\alpha$  atoms. The brassinolide molecule is sandwiched between the insertion domain and the concave surface of the solenoid structure (Fig. 1c).

### Interruption of BRI1(LRR)s by a folded domain

Two short segments (residues 590–595 and 644–646) from the insertion domain do not have interpretable electron density in the structure of brassinolide-free BRI1(LRR) but become well defined after brassinolide binding (Supplementary Fig. 3). Residues 596–643 form a separate folded domain that consists of one three-stranded antiparallel  $\beta$  sheet and one  $3_{10}$  helix (Fig. 2a). In addition, this domain contains one disulphide bridge (Cys 609–Cys 635) that can further contribute to its structural integrity. A database search using the Dali server ([http://ekhidna.biocenter.helsinki.fi/dali\\_server](http://ekhidna.biocenter.helsinki.fi/dali_server)) did not reveal known structures that share significant homology with the insertion domain, indicating that it represents a novel fold.

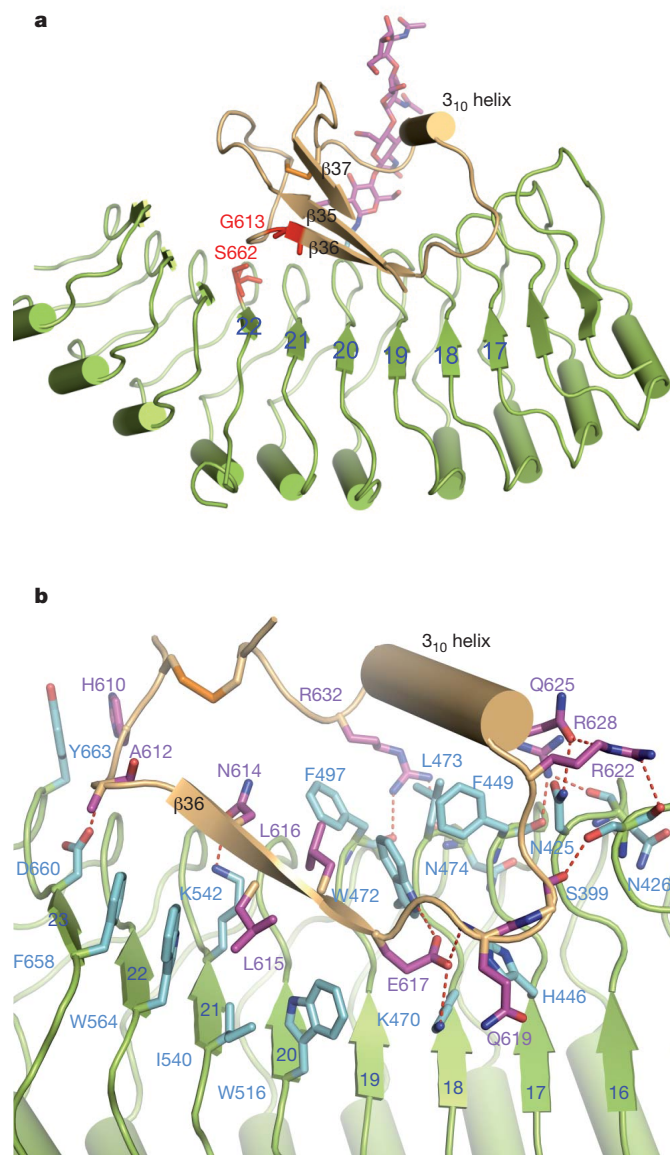
Extensive non-covalent interactions including van der Waals and polar interactions exist between the insertion domain and the solenoid structure. While making few contacts with the remaining parts of the

The blue numbers in **a** indicate the positions of LRRs. The  $\beta$  strand ( $\beta 1$ ) from the N-terminal cap is labelled. **c**, Overall structure of brassinolide-bound BRI1(LRR) with the same orientation as **a**. Brassinolide molecule is shown in stick representation and coloured in yellow.

insertion domain, the  $3_{10}$  helix packs tightly against Phe 449 and Leu 473 from underneath and forms a large network of hydrogen bonding interactions with those from one flanking side (Fig. 2b). In contrast with the  $3_{10}$  helix,  $\beta 36$  primarily binds the concave surface (Fig. 2a). Around this binding site, Trp 516, Ile 540, Trp 564 and Phe 658 establish close contacts with one side of  $\beta 36$ , whereas Trp 472 and Phe 497 wedge between  $\beta 36$  and the  $3_{10}$  helix (Fig. 2b). Hydrogen bonds dominate the interactions of the loop linking the  $3_{10}$  helix and  $\beta 36$  with the concave surface. Additionally, packing of the carbohydrate moiety of the glycosylated Asn 545 against the loop connecting the  $3_{10}$  helix and  $\beta 37$  seems to have a role in positioning the insertion domain in the concave surface (Fig. 2a). Two mutations, G613S and S662F, generated weak hormone-insensitive phenotypes<sup>2</sup>. Our structure indicates that they can perturb local conformations and consequently generate a deleterious effect on BRI1 recognition of brassinosteroids. Substitution of Gly 613 with serine would produce steric clash with the carbonyl oxygen of Ile 600 or the benzene ring of Tyr 599 and consequently generate a damaging effect on the  $\beta$  sheet of the insertion domain (Supplementary Fig. 4). Ser 662 is limited within a small pocket and its mutation to the bulky residue phenylalanine is expected to generate serious clash with its neighbouring residues, in particular the Gly 611 that is N terminal to  $\beta 36$ .

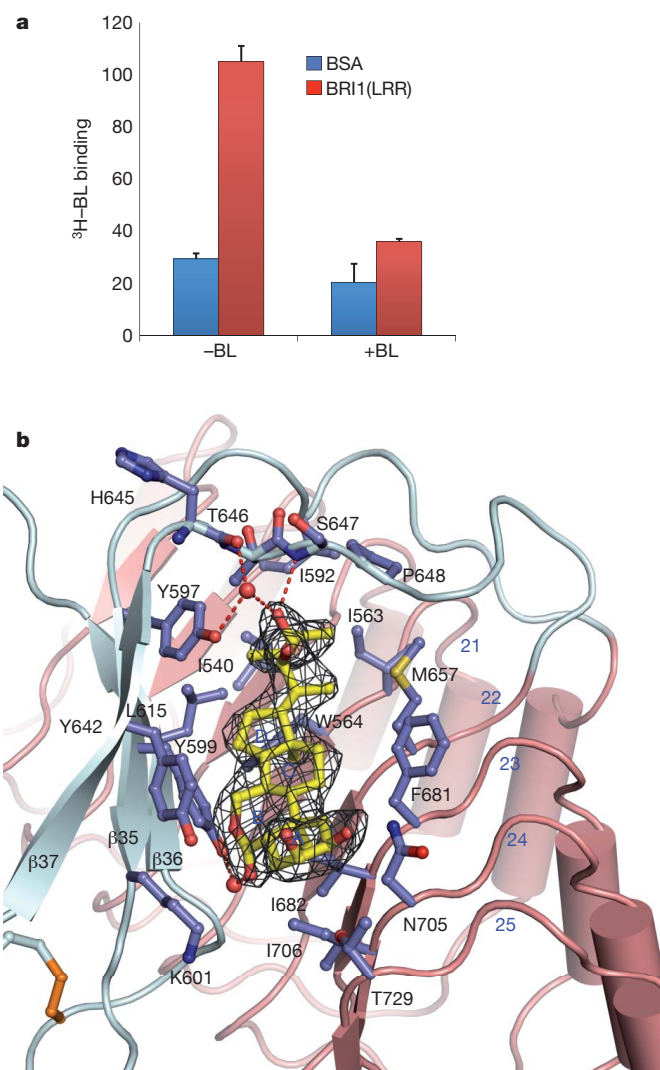
### Brassinolide binds to a surface groove on BRI1(LRR)

We used  $^3\text{H}$ -labelled brassinolide to test if the purified recombinant BRI1(LRR) was active in binding brassinolide. As shown in Fig. 3a, the protein showed a comparable brassinolide-binding activity to that



**Figure 2 | Interaction of the insertion domain with LRRs.** **a**, Cartoon representation of BRI1(LRR) highlighting interaction of the insertion domain with the concave surface. The positions of two genetic mutants are shown in red and stick representation. **b**, Detailed interactions of the insertion domain with LRRs. The side chains from the insertion domain are shown and labelled in purple, and those from LRRs in cyan. Red dashed lines represent hydrogen bonds.

previously reported<sup>12</sup>. In the brassinolide-bound BRI1(LRR) structure, the electron density unambiguously defines the brassinolide molecule that binds to a pronounced hydrophobic groove between the insertion domain and the concave side of the solenoid (Fig. 3b). Most of the residues lining the surface groove are hydrophobic and highly conserved (Supplementary Fig. 1). The fused ring moiety of brassinolide (Supplementary Fig. 5) occupies most of the surface groove (Fig. 3b). The A ring (see Supplementary Fig. 5) makes marginal contacts with the concave surface, whereas the B ring tightly stacks Tyr 642 from one lateral side. The other two rings are sandwiched between Phe 681 and Tyr 599. Despite a smaller size, the two methyl groups C18 and C19 form much denser hydrophobic interactions with BRI1 by docking into two cavities at the bottom. In contrast, the other side of the fused ring is nearly solvent exposed (Fig. 3b). Although extensive shape complementarities exist between brassinolide and the base of the surface groove, a water molecule fills the cavity that is sealed from the solvent region by Lys 601, bridging a



**Figure 3 | Brassinolide binds a hydrophobic groove between the insertion domain and the inner surface of LRRs.** **a**, <sup>3</sup>H-brassinolide (<sup>3</sup>H-BL) binding activity of BRI1(LRR). About 1 mg ml<sup>-1</sup> BRI1(LRR)-His (red bar) or bovine serum albumin (BSA) as control (blue bar) was incubated with 20 nM <sup>3</sup>H-brassinolide in the absence (-BL) or presence (+BL) of 20 μM unlabelled brassinolide. BRI1-bound <sup>3</sup>H-brassinolide was recovered using nickel beads and quantified by scintillation counting. Data represent the average of triplicate assays and error bars are standard deviations. **b**, Detailed interactions between brassinolide and BRI1(LRR). Shown in mesh is omit electron density (2.5σ) around brassinolide. The insertion domain and LRRs are coloured in slate grey and salmon pink, respectively. The side chains from both the insertion domain and LRRs are shown in slate grey. Red spheres represent oxygen atoms of water molecules. The three β strands from the insertion domain are labelled.

hydrogen bond between the carbonyl oxygen at C6 and the hydroxyl group of Tyr 599. Standing in contrast with the fused ring, the distal carbon atoms C24–C28 of the side chain are completely buried into a hydrophobic pocket (Fig. 3b). Further reinforcing the interactions around this contact interface, the hydroxyl group at C23 establishes a hydrogen bond with the backbone nitrogen of Ser 647 and two water-mediated hydrogen bonds with the carbonyl oxygen atoms of His 645 and the hydroxyl group of Tyr 597.

### Induced-fit brassinolide recognition by BRI1(LRR)

The crystals of both free and brassinolide-bound BRI1(LRR)s contain one protein molecule per asymmetric unit and have nearly identical packing. This indicates that homodimers formed by two crystallographic symmetry-related molecules (Supplementary Fig. 6) result from crystal packing rather than brassinolide binding. Consistently,

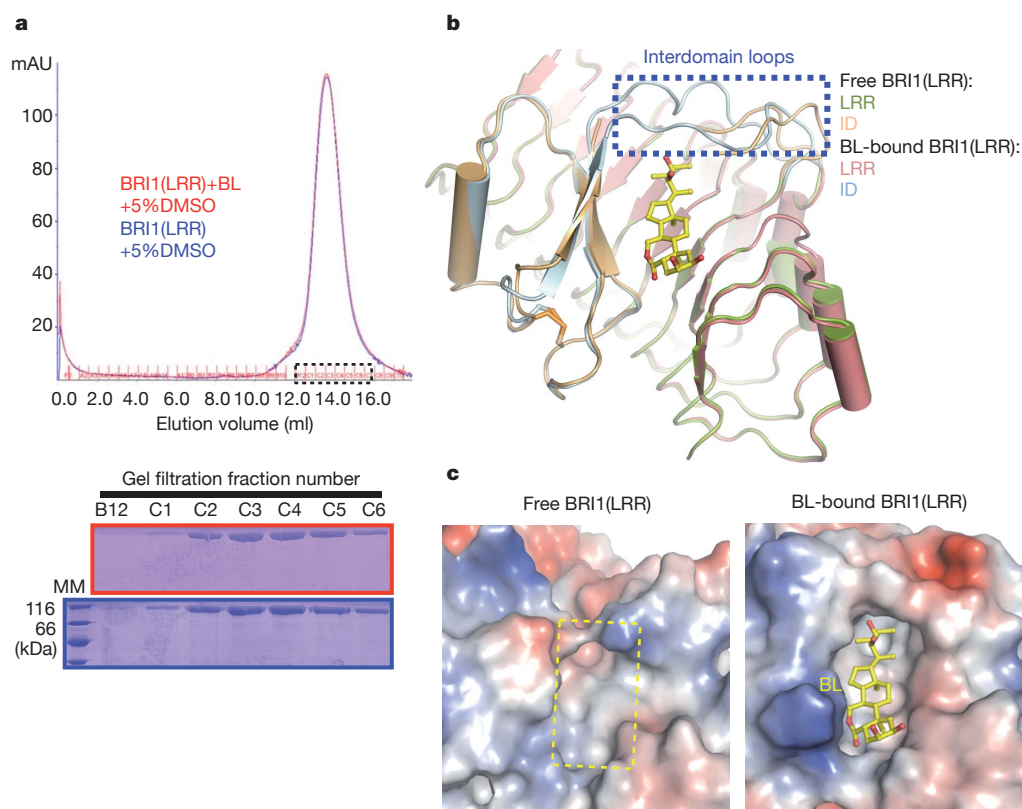


gel filtration assays showed that BRI1(LRR) was monomeric in solution and its apparent molecular weight was not affected by the presence of brassinolide (Fig. 4a). The latter is consistent with the observation<sup>22</sup> that brassinolide had no effect on BRI1 homodimerization in protoplasts. Structural comparison revealed that upon brassinolide binding marked local structural rearrangement occurs to the two loops linking the insertion domain and the LRR structure (referred to as interdomain loops) (Fig. 4b). Brassinolide-induced stabilization of the interdomain loops results in formation of a more marked surface groove where brassinolide binds (Fig. 4c), showing that brassinolide recognition by BRI1(LRR) is a process of induced fit.

## Discussion

The hydrophobicity-dominating brassinolide-binding surface groove (Fig. 4c, right) indicates that BRI1(LRR) may be accommodating in ligand binding. This would afford an explanation for why BRI1 is able to respond to a variety of brassinolide derivatives<sup>23</sup>. BRI1 ligand selectivity, on the other hand, can be conferred by the distal polar groups involved in hydrogen bonding interactions as well as the actual shape of the hormone-binding groove. The structural mechanism of brassinolide recognition by BRI1(LRR) rationalizes (Supplementary Fig. 7) the data on the structure–activity relationship of brassinosteroids accumulated during the past decades<sup>23</sup>, thus opening new perspectives for designing and developing nonsteroidal mimetic of brassinosteroids with low cost that can be widely applied in agricultural practice.

Our structural analyses reveal that, although the overall structure of BRI1(LRR) remains nearly unchanged upon brassinolide binding, significant structural rearrangement occurs to the interdomain loops around the brassinolide-binding site (collectively referred to as the brassinolide-created surface), which should be associated with brassinosteroid-induced signalling. It is not clear how this region contributes to brassinosteroid-initiated signalling, but it is probably involved in stabilization<sup>16</sup> of the kinase-domain-mediated BRI1 homodimers<sup>19</sup>. The absence of an effect of brassinolide on BRI1 homodimerization in the current study may be due to the isolated ectodomain used for assays or requirement of another protein or peptide, as indicated by the suppression of *bri1* mutants by a secreted peptidase<sup>24</sup>. If this is the case, BRI1 would be different from Toll-like receptors, in which ligand-induced homodimerization for activation can be recapitulated through their isolated extracellular domains<sup>25,26</sup>. Stabilization of BRI1 homodimers by the brassinolide-created surface is expected to initially activate BRI1 kinase to a basal level, resulting in tyrosine phosphorylation and disassociation of BKI1 (ref. 19); disassociation of BKI1 allows stable association and sequential reciprocal phosphorylation between the kinase domains of BRI1 and BAK1, fully activating the BRI1 kinase as proposed by the sequential phosphorylation model<sup>18</sup>. Alternatively, brassinolide binding may initiate signalling by altering an interaction between the extracellular domains of BRI1 and BAK1. A role of the extracellular domain of BAK1 in brassinosteroid signalling has been indicated by the observations that mutations in this domain of BAK1 affect brassinosteroid sensitivity<sup>27</sup>.



**Figure 4 | Brassinolide induces stabilization of two interdomain loops but no dimerization of BRI1(LRR).** **a**, Brassinolide (BL) has no effect on the oligomeric status of BRI1(LRR) in solution. Top, superposition of the gel filtration chromatograms of BRI1(LRR) in the absence (blue) and presence (red) of brassinolide. The vertical and horizontal axes represent ultraviolet absorbance ( $\lambda = 280$  nm) and elution volume, respectively. Peak fractions are highlighted within the dashed square. The apparent molecular weight of BRI1(LRR) was 109.4 kDa in the presence or absence of brassinolide, higher than the theoretical BRI1(LRR) monomer (83 kDa), probably owing to the existence of multiple glycosylation sites in BRI1(LRR). DMSO, dimethylsulphoxide. Bottom, Coomassie blue staining of the peak fractions

shown in the top of the panel after SDS–PAGE. MM, molecular mass marker. **b**, Brassinolide binding induces stabilization of two interdomain loops. Structural superimposition of the free and brassinolide-bound BRI1(LRR) around the brassinolide-binding site. **c**, Brassinolide binding generates a marked hydrophobic surface groove on BRI1(LRR). Shown on the left and right are the electrostatic surfaces of free BRI1(LRR) and brassinolide-bound BRI1(LRR) (shown in the same orientation) around the brassinolide-binding site, respectively. The area highlighted with the yellow dashed square on the left panel is the brassinolide-binding site. White, blue and red indicate neutral, positive and negative surfaces, respectively.



and interaction with BRI1<sup>28</sup>. Although stable BRI1–BAK1 association is a consequence of initial BRI1 activation<sup>19</sup>, a possible role of BAK1 and its homologues in initial BRI1 activation has not been excluded by genetic studies; for example, it remains unclear whether mutations of BAK1 and its homologues affect brassinolide-induced BKI1 phosphorylation by BRI1. It is interesting to note that BAK1 has five LRRs and that the new brassinolide-created surface is also located about five LRRs from the membrane surface. Thus, it is also possible that the brassinolide-created surface is involved in interaction with BAK1(LRR), which triggers BRI1 phosphorylation of BKI1 through unknown mechanisms, allowing formation of more stable BRI1–BAK1 receptor complexes. Although the brassinolide-created surface is proposed to be involved in BRI1(LRR) homodimerization or heterodimerization with BAK1(LRR), further studies are needed to determine if brassinosteroids initiate signalling through promoting protein–protein interactions, as many other plant hormones do<sup>29,30</sup>.

## METHODS SUMMARY

The extracellular domain (residues 24–784) of *Arabidopsis* BRI1 fused with 6×His at the C terminus was expressed with high five cells and a baculovirus expression system. Protein was harvested from the media and purified using Ni-NTA followed by size-exclusion chromatography (Superdex200 10/300 GL column; GE Healthcare). Crystals of free BRI1 and brassinolide-bound BRI1 were grown using the hanging-drop vapour-diffusion method and their structures were determined using molecular replacement. <sup>3</sup>H-labelled brassinolide binding assay was performed as previously described<sup>12</sup>.

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Received 10 February; accepted 6 May 2011.**

**Published online 12 June 2011.**

- Clouse, S. D. & Sasse, J. M. BRASSINOSTEROIDS: essential regulators of plant growth and development. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **49**, 427–451 (1998).
- Vert, G., Nemhauser, J. L., Geldner, N., Hong, F. & Chory, J. Molecular mechanisms of steroid hormone signaling in plants. *Annu. Rev. Cell Dev. Biol.* **21**, 177–201 (2005).
- Li, J. & Chory, J. A putative leucine-rich repeat receptor kinase involved in brassinosteroid signal transduction. *Cell* **90**, 929–938 (1997).
- Li, J. *et al.* BAK1, an *Arabidopsis* LRR receptor-like protein kinase, interacts with BRI1 and modulates brassinosteroid signaling. *Cell* **110**, 213–222 (2002).
- Nam, K. H. & Li, J. BRI1/BAK1, a receptor kinase pair mediating brassinosteroid signaling. *Cell* **110**, 203–212 (2002).
- Li, J. & Nam, K. H. Regulation of brassinosteroid signaling by a GSK3/SHAGGY-like kinase. *Science* **295**, 1299–1301 (2002).
- Sun, Y. *et al.* Integration of brassinosteroid signal transduction with the transcription network for plant growth regulation in *Arabidopsis*. *Dev. Cell* **19**, 765–777 (2010).
- Wang, X. & Chory, J. Brassinosteroids regulate dissociation of BKI1, a negative regulator of BRI1 signaling, from the plasma membrane. *Science* **313**, 1118–1122 (2006).
- Tang, W. *et al.* BSKs mediate signal transduction from the receptor kinase BRI1 in *Arabidopsis*. *Science* **321**, 557–560 (2008).
- Shiu, S. H. *et al.* Comparative analysis of the receptor-like kinase family in *Arabidopsis* and rice. *Plant Cell* **16**, 1220–1234 (2004).
- He, Z. *et al.* Perception of brassinosteroids by the extracellular domain of the receptor kinase BRI1. *Science* **288**, 2360–2363 (2000).
- Kinoshita, T. *et al.* Binding of brassinosteroids to the extracellular domain of plant receptor kinase BRI1. *Nature* **433**, 167–171 (2005).
- Li, J. Brassinosteroid signaling: from receptor kinases to transcription factors. *Curr. Opin. Plant Biol.* **8**, 526–531 (2005).
- Kim, T. W. & Wang, Z. Y. Brassinosteroid signal transduction from receptor kinases to transcription factors. *Annu. Rev. Plant Biol.* **61**, 681–704 (2010).
- Russinova, E. *et al.* Heterodimerization and endocytosis of *Arabidopsis* brassinosteroid receptors BRI1 and AtSERK3 (BAK1). *Plant Cell* **16**, 3216–3229 (2004).
- Wang, X. *et al.* Autoregulation and homodimerization are involved in the activation of the plant steroid receptor BRI1. *Dev. Cell* **8**, 855–865 (2005).
- Wang, X. *et al.* Identification and functional analysis of *in vivo* phosphorylation sites of the *Arabidopsis* BRASSINOSTEROID-INSENSITIVE1 receptor kinase. *Plant Cell* **17**, 1685–1703 (2005).
- Wang, X. *et al.* Sequential transphosphorylation of the BRI1/BAK1 receptor kinase complex impacts early events in brassinosteroid signaling. *Dev. Cell* **15**, 220–235 (2008).
- Jailais, Y. *et al.* Tyrosine phosphorylation controls brassinosteroid receptor activation by triggering membrane release of its kinase inhibitor. *Genes Dev.* **25**, 232–237 (2011).
- Kajava, A. V. Structural diversity of leucine-rich repeat proteins. *J. Mol. Biol.* **277**, 519–527 (1998).
- Di Matteo, A. *et al.* The crystal structure of polygalacturonase-inhibiting protein (PGIP), a leucine-rich repeat protein involved in plant defense. *Proc. Natl Acad. Sci. USA* **100**, 10124–10128 (2003).
- Hink, M. A., Shah, K., Russinova, E., de Vries, S. C. & Visser, A. J. Fluorescence fluctuation analysis of *Arabidopsis thaliana* somatic embryogenesis receptor-like kinase and brassinosteroid insensitive 1 receptor oligomerization. *Biophys. J.* **94**, 1052–1062 (2008).
- Zullo, M. T. A. & Adam, G. Brassinosteroid phytohormones—structure, bioactivity and applications. *Braz. J. Plant Physiol.* **14**, 143–181 (2002).
- Li, J., Lease, K. A., Tax, F. E. & Walker, J. C. BRS1, a serine carboxypeptidase, regulates BRI1 signaling in *Arabidopsis thaliana*. *Proc. Natl Acad. Sci. USA* **98**, 5916–5921 (2001).
- Liu, L. *et al.* Structural basis of Toll-like receptor 3 signaling with double-stranded RNA. *Science* **320**, 379–381 (2008).
- Park, B. S. *et al.* The structural basis of lipopolysaccharide recognition by the TLR4–MD-2 complex. *Nature* **458**, 1191–1195 (2009).
- Whippo, C. W. & Hangarter, R. P. A brassinosteroid-hypersensitive mutant of BAK1 indicates that a convergence of photomorphogenic and hormonal signaling modulates phototropism. *Plant Physiol.* **139**, 448–457 (2005).
- Yun, H. S. *et al.* Analysis of phosphorylation of the BRI1/BAK1 complex in *Arabidopsis* reveals amino acid residues critical for receptor formation and activation of BR signaling. *Mol. Cells* **27**, 183–190 (2009).
- Sheard, L. B. & Zheng, N. Plant biology: signal advance for abscisic acid. *Nature* **462**, 575–576 (2009).
- Sheard, L. B. *et al.* Jasmonate perception by inositol-phosphate-potentiated COI1–JAZ co-receptor. *Nature* **468**, 400–405 (2010).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank S. Huang and J. He at Shanghai Synchrotron Radiation Facility (SSRF) for assistance with data collection; J. Chory from the Salk Institute for providing the tritium-labelled brassinolide. This research was funded by the National Outstanding Young Scholar Science Foundation of National Natural Science Foundation of China grant no. 20101331722 to J. Chai and NIH R01GM066258 to Z.-Y.W.

**Author Contributions** J. Chai, Z.H., J.S. and Z.-Y.W. designed the experiments. The binding assay was performed by T.-W.K. and the other assays by J.S., Z.H., Jinjing W., W.C., Jiawei W., M.Y., S.S. and J. Chang. Data were analysed by J. Chai, Z.H., J.S. and Z.-Y.W. J. Chai, Z.-Y.W. and Z.H. wrote the paper.

**Author Information** The atomic coordinates and structure factors of BRI1(LRR) and the BRI1(LRR)–brassinolide complex have been deposited in the Protein Data Bank under the accession codes 3RGX and 3RGZ, respectively. Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at [www.nature.com/nature](http://www.nature.com/nature). Correspondence and requests for materials should be addressed to J. Chai ([chaijj@mail.tsinghua.edu.cn](mailto:chaijj@mail.tsinghua.edu.cn)).

## METHODS

**Protein expression and purification.** The LRR domain of BRI1 (residues 24–784) from *Arabidopsis* with an engineered C-terminal 6×His tag was generated by standard PCR-based cloning strategy and its identity was confirmed by sequencing. The protein was expressed in high five cells using the vector pFastBac 1 (Invitrogen) with a modified N-terminal Hemolin peptide. One litre of cells ( $2.0 \times 10^6$  cells ml<sup>-1</sup>) was infected with 20 ml baculovirus using a multiplicity of infection of 4 at 22 °C, and protein was harvested from the media after 48 h. The protein was purified using Ni-NTA (Novagen) and size-exclusion chromatography (Superdex 200, Pharmacia) in buffer (10 mM Tris, pH 8.0, 100 mM NaCl). Samples from relevant fractions were applied to SDS-PAGE and visualized by Coomassie blue staining. Protein purification was performed at 4 °C. For crystallization of BRI1(LRR), the purified protein was concentrated to about 3.0 mg ml<sup>-1</sup> in buffer containing 10 mM Tris, pH 8.0, 100 mM NaCl.

**Crystallization, data collection, structure determination and refinement.** Crystals of BRI1(LRR) were generated by mixing the protein with an equal amount of well solution (1 µl) by the hanging-drop vapour-diffusion method. A mixture of BRI1(LRR) and brassinolide with a molar ratio of 1:10 was used for generating crystals of their complex. The initial buffer producing crystals of both free BRI1(LRR) and brassinolide-bound BRI1(LRR) contained 0.2 M Na<sub>2</sub>SO<sub>4</sub> and 20% (w/v) polyethylene glycol (PEG) 3,350, which was further optimized by adding 10 mM trimethylamine-HCl. Crystals grew to their maximum size ( $0.1 \times 0.1 \times 0.1$  mm<sup>3</sup>) within 10 days at room temperature (20 °C).

All the diffraction data sets were collected at the Shanghai Synchrotron Radiation Facility (SSRF) at beam line BL17U1 using a CCD detector. Crystals of both brassinolide-free and brassinolide-bound BRI1(LRR) belong to space group C2 with one protein molecule per asymmetric unit. For data collection, the crystals were equilibrated in a cryoprotectant buffer containing reservoir buffer plus 20.0% (v/v) glycerol. The data were processed using HKL2000 (ref. 31). Molecular replacement (MR) with the program PHASER<sup>32</sup> was used to solve the crystal structures of both brassinolide-free and brassinolide-bound BRI1(LRR). The atomic coordinates of PGIP2 (ref. 21) (PDB code 1OGQ) and InIA<sup>33</sup> (PDB code 1O6V) were used as the initial searching model. The model from MR was built with the program COOT<sup>34</sup> and subsequently subjected to refinement by the program PHENIX<sup>35</sup>. The electron density for brassinolide in BRI1(LRR)–brassinolide complex crystal became apparent after refinement of BRI1(LRR). The final refined model contains residues 30–775 of BRI1(LRR) and one brassinolide molecule in brassinolide-bound BRI1(LRR), and residues 30–772 in brassinolide-free BRI1(LRR) except the two segments containing residues 590–595 and 644–646 that have no clear electron density and are presumed to be disordered in solution. The structure figures were prepared using PyMOL<sup>36</sup>.

**Gel filtration assay.** BRI1(LRR) protein purified as described earlier was subjected to gel filtration analysis (Superdex200 10/300 GL column; GE Healthcare) in the presence and absence of brassinolide. Buffer containing 50 mM NaH<sub>2</sub>PO<sub>4</sub>/Na<sub>2</sub>HPO<sub>4</sub> pH 7.4, 100 mM NaCl, 2.0 mM MgCl<sub>2</sub> and 5% (v/v) DMSO was used for the assay. A mixture of BRI1(LRR) and brassinolide with a molar ratio of about 1:10 was used to test the effect of brassinolide on the apparent molecular weight of BRI1(LRR). The assays were performed with a flow rate of 0.5 ml min<sup>-1</sup> and an injection volume of 0.5 ml buffer containing BRI1(LRR) (0.83 mg ml<sup>-1</sup>) with or without brassinolide at 20 °C. The column was calibrated using five standard proteins, β-amylase 200 kDa, alcohol dehydrogenase 150 kDa, albumin 66 kDa, carbonic anhydrase 29 kDa and cytochrome C 12.4 kDa, which had elution volumes ( $V_e$ ) of 12.53 ml, 13.25 ml, 14.25 ml, 16.18 ml and 16.97 ml, respectively, under the conditions of assay. The void volume ( $V_o$ ) for the column used was determined to be 8.15 ml using fresh blue dextran solution (1.0 mg ml<sup>-1</sup>). The calibration curve of the gel-phase distribution coefficient ( $K_{av}$ ) versus log molecular weight ( $M$ ) was obtained using Excel.  $K_{av} = (V_e - V_o)/(V_c - V_o)$ , where  $V_e$  is the elution volume,  $V_o$  is the void volume,  $V_c$  is the geometric column volume, 24 ml for the column used. The elution volume of BRI1(LRR) in the presence and absence of brassinolide was 13.63 ml under the conditions of assay, corresponding to an apparent molecular weight of 109.4 kDa based on the calibration equation  $K_{av} = -0.2366 \lg M + 1.538$ ,  $R^2 = 0.9814$ .

**<sup>3</sup>H-labelled brassinolide-binding assay.** His-tagged BRI1(LRR), or BSA as control, was incubated with 20 nM <sup>3</sup>H-brassinolide, or 20 nM <sup>3</sup>H-brassinolide mixed with 20 µM non-radiolabelled brassinolide as competitor, and nickel beads in the brassinolide-binding buffer (0.25 M mannitol, 10 mM Tris-2-[N-morpholino]ethanesulphonic acid (MES), pH 5.7, 5 mM MgCl<sub>2</sub>, 0.1 mM CaCl<sub>2</sub>) for 30 min. The beads were washed three times, and His-BRI1(LRR) and bound [<sup>3</sup>H]-brassinolide were eluted by using 1 M imidazole. The eluted radioactivity was measured by scintillation counter.

31. Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326 (1997).
32. McCoy, A. J. *et al.* Phaser crystallographic software. *J. Appl. Cryst.* **40**, 658–674 (2007).
33. Schubert, W. D. *et al.* Structure of internalin, a major invasion protein of *Listeria monocytogenes*, in complex with its human receptor E-cadherin. *Cell* **111**, 825–836 (2002).
34. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **60**, 2126–2132 (2004).
35. Adams, P. D. *et al.* PHENIX: building new software for automated crystallographic structure determination. *Acta Crystallogr. D* **58**, 1948–1954 (2002).
36. DeLano, W. L. PyMOL molecular viewer (<http://www.pymol.org>) (2002).

# Subunit arrangement and phenylethanolamine binding in GluN1/GluN2B NMDA receptors

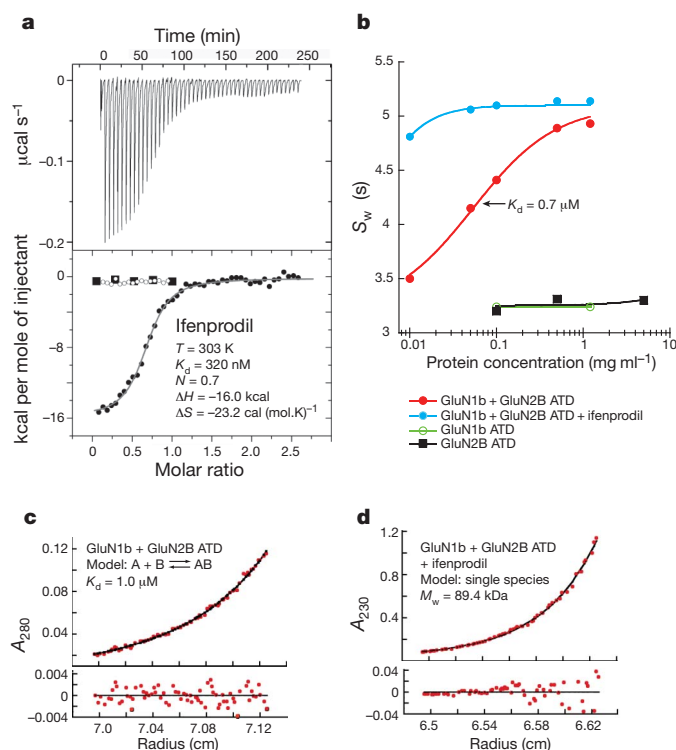
Erkan Karakas<sup>1</sup>, Noriko Simorowski<sup>1</sup> & Hiro Furukawa<sup>1</sup>

Since it was discovered that the anti-hypertensive agent ifenprodil has neuroprotective activity through its effects on NMDA (N-methyl-D-aspartate) receptors<sup>1</sup>, a determined effort has been made to understand the mechanism of action and to develop improved therapeutic compounds on the basis of this knowledge<sup>2–4</sup>. Neurotransmission mediated by NMDA receptors is essential for basic brain development and function<sup>5</sup>. These receptors form heteromeric ion channels and become activated after concurrent binding of glycine and glutamate to the GluN1 and GluN2 subunits, respectively. A functional hallmark of NMDA receptors is that their ion-channel activity is allosterically regulated by binding of small compounds to the amino-terminal domain (ATD) in a subtype-specific manner. Ifenprodil and related phenylethanolamine compounds, which specifically inhibit GluN1 and GluN2B NMDA receptors<sup>6,7</sup>, have been intensely studied for their potential use in the treatment of various neurological disorders and diseases, including depression, Alzheimer's disease and Parkinson's disease<sup>2,4</sup>. Despite considerable enthusiasm, mechanisms underlying the recognition of phenylethanolamines and ATD-mediated allosteric inhibition remain limited owing to a lack of structural information. Here we report that the GluN1 and GluN2B ATDs form a heterodimer and that phenylethanolamine binds at the interface between GluN1 and GluN2B, rather than within the GluN2B cleft. The crystal structure of the heterodimer formed between the GluN1b ATD from *Xenopus laevis* and the GluN2B ATD from *Rattus norvegicus* shows a highly distinct pattern of subunit arrangement that is different from the arrangements observed in homodimeric non-NMDA receptors and reveals the molecular determinants for phenylethanolamine binding. Restriction of domain movement in the bi-lobed structure of the GluN2B ATD, by engineering of an inter-subunit disulphide bond, markedly decreases sensitivity to ifenprodil, indicating that conformational freedom in the GluN2B ATD is essential for ifenprodil-mediated allosteric inhibition of NMDA receptors. These findings pave the way for improving the design of subtype-specific compounds with therapeutic value for neurological disorders and diseases.

The consensus view that has emerged from functional studies of NMDA receptors using site-directed mutagenesis and molecular modelling is that phenylethanolamine compounds such as ifenprodil and Ro 25-6981 bind to the ATD of the GluN2B subunit. However, this has not been established directly and the mechanism of action is complicated by the obligate heteromeric assembly of NMDA receptors. To establish directly that phenylethanolamines bind to the ATDs of these receptors, we used isothermal titration calorimetry to measure the binding of ifenprodil and Ro 25-6981 to purified recombinant *Rattus norvegicus* GluN2B (residues 31–394) and *Xenopus laevis* GluN1b (residues 23–408) ATDs (Supplementary Fig. 1). GluN1b from *Xenopus laevis*<sup>8,9</sup> was used in this study because of its superior biochemical stability compared to other orthologues. It is 93% identical in primary sequence to the *Rattus norvegicus* GluN1 ATD and is capable of forming functional NMDA-receptor ion channels that undergo ifenprodil inhibition when combined with *Rattus norvegicus* GluN2B<sup>9</sup> (Supplementary Fig. 2).

When the GluN1b ATD or GluN2B ATD proteins were individually injected with ifenprodil, there was no evidence of binding (Fig. 1a). However, when a mixture of the GluN1b and GluN2B ATD proteins was injected with ifenprodil or Ro 25-6981, a dose-dependent heat exchange was observed, with dissociation constant ( $K_d$ ) values of 320 nM and 60 nM, respectively (Fig. 1a and Supplementary Fig. 3). Thus, both the GluN1b and GluN2B ATDs are required for binding of phenylethanolamines.

The necessity of both ATDs for recognition of phenylethanolamine indicates that binding takes place in the GluN1–GluN2B heteromer. To probe the association pattern of GluN1b and GluN2B ATD proteins,



**Figure 1 | Binding of phenylethanolamine requires both GluN1b and GluN2B ATDs, and stabilizes heterodimers.** **a**, Calorimetric titration of ifenprodil into a GluN1b and GluN2B ATD mixture (upper panel) and integrated heat as a function of the ifenprodil/protein molar ratio (lower panel) for GluN1b ATD (open circles), GluN2B ATD (filled squares) and the GluN1b/GluN2B ATD mixture (filled circles). **b**, Weighted-average sedimentation coefficient ( $S_w$ ) for GluN1b ATD alone (green), GluN2B ATD alone (black) and the GluN1b–GluN2B ATD mixture in the presence (cyan) and absence (red) of 10  $\mu\text{M}$  ifenprodil, fitted with a monomer-dimer model (lines). **c**, **d**, Sedimentation equilibrium analysis of GluN1b and GluN2B ATDs in the absence (**c**) and presence (**d**) of 10  $\mu\text{M}$  ifenprodil. Data points at a rotor speed of 18,000 r.p.m. (red dots) are shown with a global fit (black line) of the data. Residuals from the fit are shown in the lower panel.

<sup>1</sup>Cold Spring Harbor Laboratory, WM Keck Structural Biology Laboratory, 1 Bungtown Road, Cold Spring Harbor, New York 11724, USA.

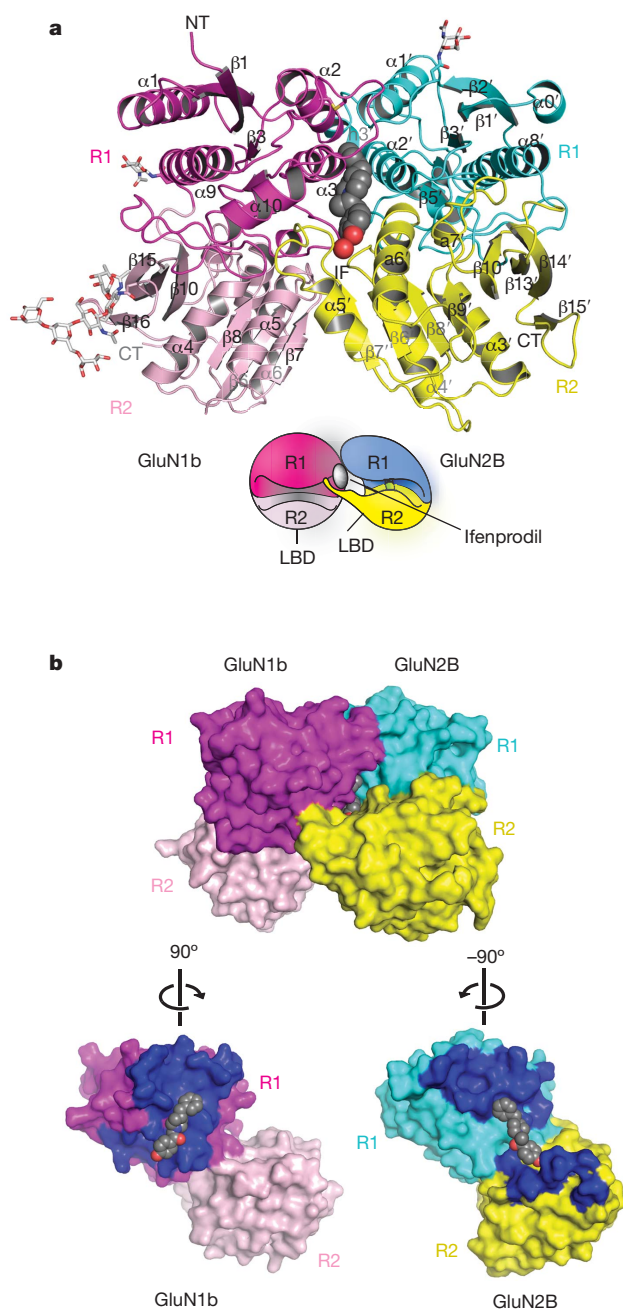


we determined the mass of the ATD proteins in solution by sedimentation experiments (Fig. 1b–d). Although the individual GluN1b ATD and GluN2B ATD were exclusively monomeric at  $1.2 \text{ mg ml}^{-1}$  (Fig. 1b), they formed a heterodimer with a  $K_d$  of  $0.7\text{--}1 \mu\text{M}$  when mixed together (Fig. 1b, c). Notably, when ifenprodil was included in the GluN1b/GluN2B ATD protein mixture, the heterodimerization was strengthened by at least 20-fold (Fig. 1b, d). These results establish that the GluN1b and GluN2B ATDs form heterodimers and that phenylethanolamines probably bind at the GluN1b–GluN2B subunit interface.

To understand the nature of the subunit interaction between GluN1b and GluN2B at their ATDs, and to pinpoint the location of the phenylethanolamine binding site, we conducted crystallographic studies on the GluN1b and GluN2B ATD proteins (Supplementary Table 1). The crystallographic analysis showed that the GluN1b and GluN2B ATDs exist as heterodimers in both ifenprodil-bound and Ro 25-6981-bound forms (Fig. 2). No notable structural difference was observed between the monomers of GluN1b ATD (Supplementary Fig. 4) or GluN2B ATD<sup>10</sup> and the respective subunits in the GluN1b–GluN2B ATD complex, indicating that dimerization did not cause changes in the overall conformation. Most notably, the crystal structures clearly identified the phenylethanolamine binding site at the heterodimer interface (Fig. 2).

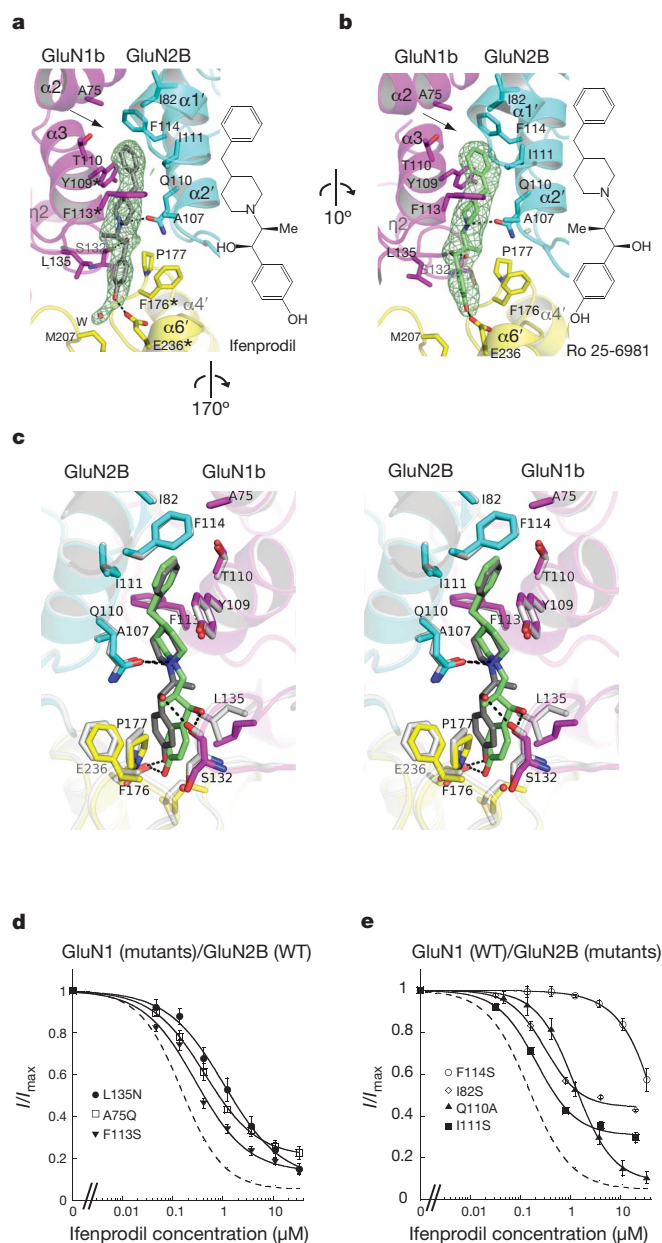
Both the GluN1b and GluN2B ATDs have bi-lobed clamshell-like architectures composed of R1 and R2 domains that are roughly similar in secondary-structure distribution to non-NMDA-receptor ATDs<sup>11–14</sup>. However, the structures of the GluN1b and GluN2B ATD monomers cannot be superimposed onto non-NMDA-receptor ATD monomers, owing to a major difference in the R1–R2 orientations, as was also observed previously in a study of the GluN2B ATD monomer<sup>10</sup> (Supplementary Fig. 5). The unique R1–R2 orientations of the GluN1b and GluN2B ATDs result in a heterodimer assembly that is distinct from that observed in non-NMDA-receptor ATD homodimers<sup>11–14</sup> (Supplementary Fig. 6). Whereas non-NMDA-receptor ATD subunits form symmetrical homodimers through strong R1–R1 and R2–R2 interactions, the GluN1b and GluN2B ATDs associate with each other asymmetrically through R1–R1 and R1(GluN1b)–R2(GluN2B) interactions<sup>11,12</sup> (Fig. 2). No residue from GluN1b R2 is involved in the GluN1b–GluN2B interaction (Fig. 2b). The R1–R1 interface contains hydrophobic interactions mediated by residues from the cores of the  $\alpha 2$  helix and  $\alpha 3$  helix in GluN1b, and from the  $\alpha 1'$  helix and  $\alpha 2'$  helix in GluN2B, surrounded by polar interactions involving the GluN1b  $\alpha 2$  helix, the GluN2B  $\alpha 1'$  helix and the hypervariable loops<sup>10</sup> (Supplementary Fig. 7). The R1–R2 interface involves mainly polar interactions, involving residues on the  $\alpha 10$  helix, a loop extending from  $\eta 2$  in GluN1b and loops extending from the  $\beta 6'$  sheet and  $\beta 7'$  sheet in GluN2B (Fig. 2d). The lack of R2–R2 interaction in the GluN1b and GluN2B ATDs leaves sufficient room for the previously suggested conformational movement of the bi-lobed structure in GluN2B<sup>10,15</sup>, which is important in mediating the allosteric regulation that is unique to NMDA receptors. In non-NMDA receptors, such movement is prohibited, owing to strong R2–R2 interactions that lock the movement of R2 (refs 3, 11–13).

The heterodimeric arrangement of GluN1b and GluN2B creates a phenylethanolamine binding pocket composed of residues from GluN1b R1, GluN2B R1 and GluN2B R2 (Fig. 2). The phenylethanolamine binding site has no overlap with the zinc binding site that is located in the GluN2B ATD cleft<sup>10,16</sup> (Supplementary Fig. 8). In the crystal structure, ifenprodil is buried in the dimer interface with insufficient space for entering or exiting (Fig. 2b), which indicates that binding occurs through an induced-fit mechanism and that unbinding may involve opening of the GluN2B ATD bi-lobed structure. All of the residues at the binding sites are identical among *Xenopus laevis*, rat and human orthologues, indicating that inhibition of NMDA receptors by phenylethanolamine is a conserved feature among those species (Supplementary Fig. 9). Binding of both ifenprodil and Ro 25-6981 is



**Figure 2 | Structure of the GluN1b–GluN2B ATD heterodimer in complex with ifenprodil at 2.6 Å resolution.** **a**, View of the ATD heterodimer from the side. The GluN1b and GluN2B ATDs have bi-lobed architecture composed of R1 (magenta and cyan) and R2 (light pink and yellow) domains. Ifenprodil (grey spheres) sits at the heterodimer interface. N-glycosylation chains are shown in white. NT, N terminus; CT, C terminus. The cartoon shows an approximate orientation of the GluN1b and GluN2B ATDs with black sticks below R2 indicating the C-terminal ends where ligand-binding domains (LBDs) begin. **b**, Surface presentation of the GluN1b–GluN2B ATD heterodimer (upper panel) and of each subunit (lower panel), showing residues at the subunit interface in dark blue. Note that ifenprodil (grey spheres) is occluded in the subunit interface. The heterodimer buries  $1,191 \text{ Å}^2$  of solvent-accessible surface area per subunit, with the GluN1b R1–GluN2B R1 and GluN1b R1–GluN2B R2 interfaces contributing 62% and 38%, respectively.

mediated primarily through hydrophobic interactions between the benzylpiperidine group and a cluster of hydrophobic residues from the GluN1b  $\alpha 2$  helix and  $\alpha 3$  helix and the GluN2B  $\alpha 1'$  helix and  $\alpha 2'$  helix, and between the hydroxylphenyl groups and GluN1b Leu 135, GluN2B Phe 176 and GluN2B Pro 177 (Fig. 3a, b). Furthermore, the



**Figure 3 | Phenylethanolamine binding site.** **a, b**, Binding of ifenprodil (**a**) and Ro 25-6981 (**b**) takes place at the GluN1b–GluN2B subunit interface. Mesh represents the  $F_o - F_c$  omit electron density map contoured at  $3\sigma$ . Residues marked with asterisks in **a** have been previously shown to affect ifenprodil sensitivity. Adjacent to the binding pocket is an empty space surrounded by hydrophobic residues, including GluN1b Ala 75, GluN2B Ile 82 and GluN2B Phe 114 (arrows). **c**, Comparison of binding patterns of ifenprodil (grey) and Ro 25-6981 (lime) in stereoview. The structure bound to Ro 25-6981 is coloured as in **b**, whereas the ifenprodil-bound structure is coloured white. **d, e**, New residues found to interact with phenylethanolamines in this study were mutated and analysed for their effect on sensitivity to ifenprodil. Mutation of the residues surrounding the binding site caused changes in  $IC_{50}$  as well as changes in the extent of inhibition by ifenprodil. WT, wild-type;  $I/I_{max}$  relative current with ( $I$ ) and without ( $I_{max}$ ) ifenprodil. Error bars represent s.d.

drugs make three direct polar interactions with Ser 132 of GluN1b, Glu 110 of GluN2B and Asp 236 of GluN2B. Superposition of the binding sites of ifenprodil and Ro 25-6981 shows that the methyl and hydroxyl groups in the propanol moiety of both ligands face in opposite directions and that the benzylpiperidine groups sit in the binding pocket in similar ways (Fig. 3c). Consequently, Ro 25-6981 has a higher affinity for GluN1/GluN2B NMDA receptors than ifenprodil<sup>17</sup> because the methyl group in Ro 25-6981 is in a favourable

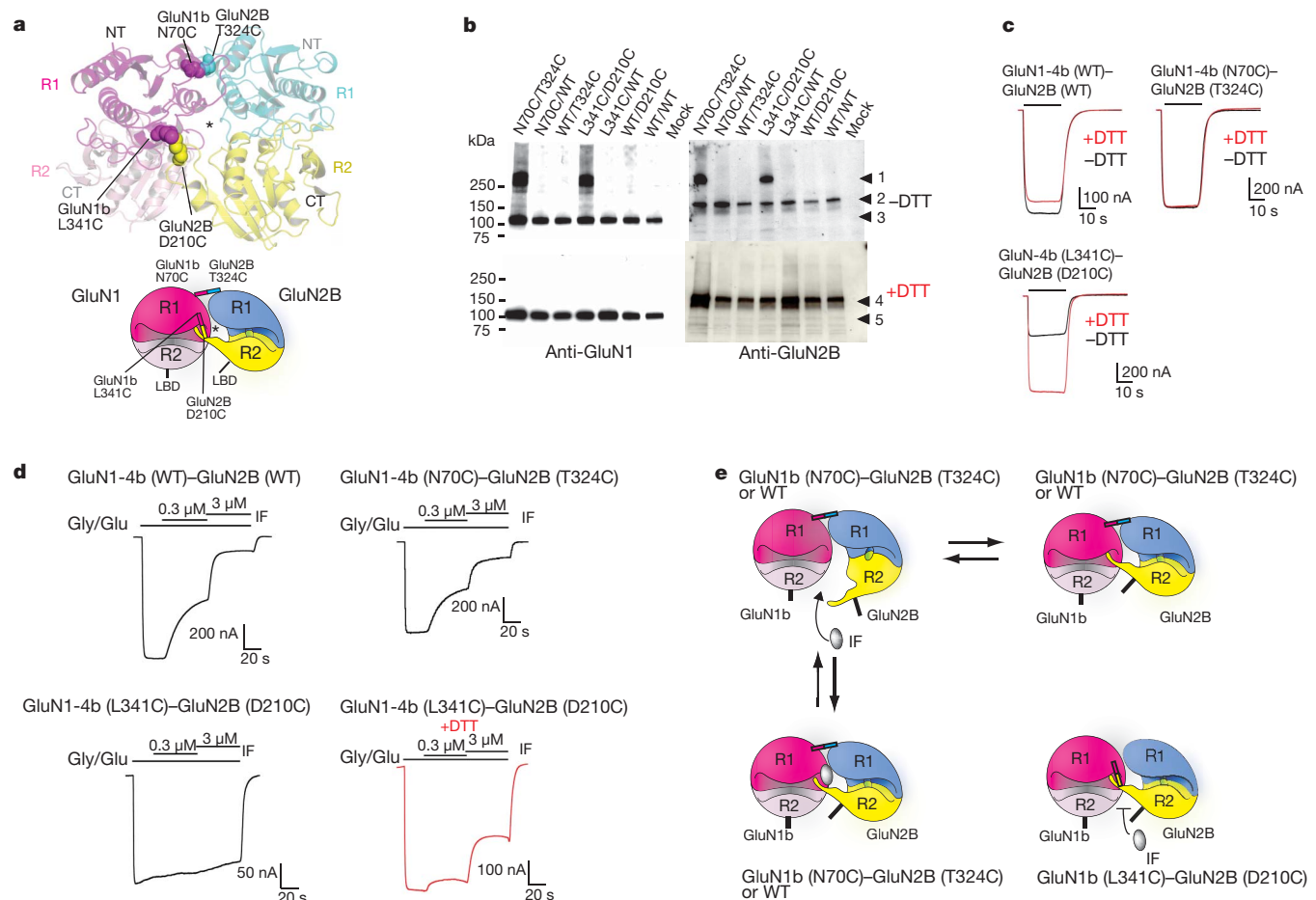
position to form a hydrophobic interaction involving Phe 176 and Pro 177 in the GluN2B subunit, whereas ifenprodil makes a weaker hydrophobic interaction with GluN1b, involving Leu 135. Extensive mutagenesis studies have previously indicated that GluN1b Tyr 109 (ref. 18) and GluN2B Phe 176 and Asp 236 (ref. 19) are critical in mediating inhibition by ifenprodil, but whether these are involved in binding or transducing the inhibitory effect was unknown. We performed additional mutagenesis studies on newly identified residues in both GluN1b and GluN2B at the ifenprodil binding site, measured macroscopic currents by two-electrode voltage clamp, and revealed significant alterations in  $IC_{50}$  and in the extent of inhibition (Fig. 3d, e and Supplementary Table 2), thereby confirming the physiological relevance of the binding site. Notably, disruption of the 'empty' hydrophobic space formed by GluN1b Ala 75, GluN2B Ile 82 and GluN2B Phe 114 (arrows in Fig. 3a and b) by site-directed mutations to hydrophilic residues had marked effects on sensitivity to ifenprodil (Fig. 3d, e). Thus, stabilization of this hydrophobic space by filling it with a hydrophobic moiety may be a valid strategy to improve the design of phenylethanolamine-based drugs.

It is not known why phenylethanolamine binds specifically to the GluN1–GluN2B subunit combination. Although inspection of the primary sequences shows non-conservation of the critical binding-site residues between GluN2B and GluN2C or GluN2D (for example, the equivalent residue to GluN2B Phe 176 is not conserved in GluN2C or GluN2D), all of the residues in GluN2A are conserved except for GluN2B Ile 111 (Met 112 in GluN2A) (Supplementary Fig. 10). Indeed, the mutations GluN2A Met112Ile or GluN2B Ile111Met do not confer or abolish ifenprodil sensitivity in GluN1/GluN2A or GluN1/GluN2B receptors, respectively (Supplementary Table 2). Thus, the insensitivity of the GluN1/GluN2A receptors to phenylethanolamine may stem from a fundamental difference in the mode of subunit association between GluN1/GluN2A and GluN1/GluN2B at their ATDs.

To validate further the physiological relevance of the heterodimeric assembly, we engineered cysteine mutants at the subunit interface, using the ifenprodil-bound GluN1b/GluN2B ATD structure as a guide, in the context of the intact rat GluN1-4b/GluN2B receptor. These cysteines were designed to form spontaneous disulphide bonds if the mutated residues were proximal to each other. We designed two pairs of cysteine mutants, GluN1-4b (Asn70Cys) with GluN2B (Thr324Cys), and GluN1-4b (Leu341Cys) with GluN2B (Asp210Cys). These mutations 'lock' the R1–R1 and R1–R2 interfaces, respectively (Fig. 4a). We then expressed the mutant receptors in mammalian cell cultures and analysed them for formation of disulphide-linked oligomers in western blots. When mutant receptors of one subunit were co-expressed with wild-type receptors of the other, they gave rise to monomeric bands that were identical to wild-type GluN1-4b–GluN2B receptors in both reducing and non-reducing conditions (110 kDa and 170 kDa for GluN1-4b and GluN2B, respectively; Fig. 4b, arrows 2 and 3). In contrast, co-expressing pairs of the GluN1-4b–GluN2B cysteine mutants gave rise to a heterodimeric ~280 kDa band that was recognized by both anti-GluN1 and anti-GluN2B antibodies in non-reducing conditions (Fig. 4b, arrow 1). This confirms that the R1–R1 and R1–R2 subunit interfaces observed in the GluN1b–GluN2B ATD crystal structures are physiological and that the heterodimer, not the homodimer, is the basic functional unit in the ATD of the NMDA receptor<sup>20</sup>. Furthermore, disulphide crosslinking was observed in the presence and absence of ifenprodil, indicating that the ligand-free GluN1b–GluN2B ATDs may oscillate between the previously suggested open conformation<sup>15</sup> and the closed conformation represented by the crystal structure described here.

To understand the functional effects of locking the R1–R1 and R1–R2 interactions in the GluN1b and GluN2B ATDs, we measured macroscopic current responses from the ion channels of the cysteine-mutant receptors by two-electrode voltage clamp. First, we explored the effect on ion-channel activity of breaking the disulphide bonds.





**Figure 4 | Engineering of disulphide bonds at the subunit interface alters sensitivity to ifenprodil.** **a**, Location of mutated residues at the R1–R1 and R1–R2 interfaces in the GluN1b and GluN2B ATDs (spheres), and location of the ifenprodil binding pocket (asterisk). **b**, Detection of disulphide bonds by anti-GluN1 and anti-GluN2B western blots in reducing (+DTT) and non-reducing (–DTT) conditions. Arrow 1, GluN1-4b–GluN2B heterodimer; arrows 2 and 4, GluN2B monomers; arrows 3 and 5, GluN1-4b monomers. **c**, Macroscopic current recording of the wild-type and mutant receptors in the presence (red)

and absence (black) of DTT (2 mM). **d**, Effect of disulphide bonds on the sensitivity to ifenprodil (IF) of wild-type and mutant receptors in the presence (red) and absence (black) of DTT. **e**, Possible model of ifenprodil binding and the movement of ATDs for allosteric inhibition. Ifenprodil binds to the open GluN2B clamshell and induces domain closure, resulting in allosteric inhibition. In the GluN1-4b (Asn70Cys)-GluN2B (Thr324Cys) receptor, the GluN2B ATD is locked in the closed conformation so ifenprodil cannot access the binding site.

Application of dithiothreitol (DTT) had a minor inhibitory effect on wild-type GluN1-4b–GluN2B receptors and on receptors containing GluN1-4b (Asn70Cys) and GluN2B (Thr324Cys). In contrast, a 2.5-fold potentiation was observed on breakage of the disulphide bond at the R1–R2 interface between GluN1-4b (Leu341Cys) and GluN2B (Asp210Cys) (Fig. 4c and Supplementary Fig. 11). This implies that locking the closed conformation in the GluN2B ATD bi-lobed structure by the R1–R2 crosslink results in downregulation of ion-channel activity. We next tested the effects of the disulphide bonds on sensitivity to ifenprodil. Although the R1–R1 crosslink had only a minor effect, the R1–R2 crosslink almost completely abolished inhibition by ifenprodil, even at 3  $\mu$ M (Fig. 4d). When this R1–R2 disulphide crosslink was broken by the application of DTT, the mutant receptors regained sensitivity to ifenprodil, to a similar extent to that of receptors composed of wild-type GluN1-4b and GluN2B (Asp210Cys) in non-reducing conditions (Fig. 4d and Supplementary Fig. 12). This indicates that ifenprodil cannot bind to the GluN1b–GluN2B ATD when the R1–R2 interaction is locked and thus, when the GluN2B ATD clamshell is closed. Taken together, the experiments described above indicate that the binding of ifenprodil requires an opening of the GluN2B bi-lobed structure and that inhibition by ifenprodil involves closure of the clamshell through the GluN1b R1–GluN2B R2 interaction (Fig. 4e).

This study shows that phenylethanolamine binds at the GluN1–GluN2B subunit interface through an induced-fit mechanism and that allosteric inhibition involves stabilization of the GluN2B ATD clamshell structure in a closed conformation. The binding mechanism presented here provides a molecular blueprint for improving the design of therapeutic compounds targeting the ATD of the NMDA receptor.

## METHODS SUMMARY

GluN1b and GluN2B ATDs were expressed as secreted proteins using the insect-cell/baculovirus system and purified using metal-chelate chromatography and size-exclusion chromatography. Crystallization was performed in hanging-drop vapour diffusion configuration in a buffer containing 20% PEG3350, 150 mM KNO<sub>3</sub> and 50 mM HEPES-NaOH (pH 7.0) for the GluN1b ATD, or 3.0–3.5 M sodium formate and 0.1 M HEPES-NaOH (pH 7.5) for the GluN1b–GluN2B ATD heterodimer. Diffraction data sets obtained at 100 K were indexed, integrated and scaled using HKL2000. The GluN1b ATD structure was solved by the single anomalous diffraction phasing method using Se-Met-incorporated crystals, and the GluN1b–GluN2B ATD structures were solved by molecular replacement using coordinates of GluN1b ATD and GluN2B ATD (Protein Data Bank code 3JPW<sup>10</sup>). Model refinement was conducted using the program Phenix<sup>21</sup>. Experiments involving analytical ultracentrifugation and isothermal titration calorimetry were conducted using the purified protein samples in their glycosylated form. Ion-channel activities of full-length NMDA receptors were measured by whole-cell recording



from cRNA-injected *Xenopus laevis* oocytes, using a two-electrode voltage-clamp configuration.

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Received 20 January; accepted 6 May 2011.**

**Published online 15 June 2011.**

- Gotti, B. *et al.* Ifenprodil and SL 82.0715 as cerebral anti-ischemic agents. I. Evidence for efficacy in models of focal cerebral ischemia. *J. Pharmacol. Exp. Ther.* **247**, 1211–1221 (1988).
- Koller, M. & Urwyler, S. Novel *N*-methyl-D-aspartate receptor antagonists: a review of compounds patented since 2006. *Expert Opin. Ther. Pat.* **20**, 1683–1702 (2010).
- Hansen, K. B., Furukawa, H. & Traynelis, S. F. Control of assembly and function of glutamate receptors by the amino-terminal domain. *Mol. Pharmacol.* **78**, 535–549 (2010).
- Mony, L., Kew, J. N., Gunthorpe, M. J. & Paoletti, P. Allosteric modulators of NR2B-containing NMDA receptors: molecular mechanisms and therapeutic potential. *Br. J. Pharmacol.* **157**, 1301–1317 (2009).
- Traynelis, S. F. *et al.* Glutamate receptor ion channels: structure, regulation, and function. *Pharmacol. Rev.* **62**, 405–496 (2010).
- Gallagher, M. J., Huang, H., Pritchett, D. B. & Lynch, D. R. Interactions between ifenprodil and the NR2B subunit of the *N*-methyl-D-aspartate receptor. *J. Biol. Chem.* **271**, 9603–9611 (1996).
- Williams, K. Ifenprodil discriminates subtypes of the *N*-methyl-D-aspartate receptor: selectivity and mechanisms at recombinant heteromeric receptors. *Mol. Pharmacol.* **44**, 851–859 (1993).
- Ewald, R. C. & Cline, H. T. Cloning and phylogenetic analysis of NMDA receptor subunits NR1, NR2A and NR2B in *Xenopus laevis* tadpoles. *Front. Mol. Neurosci.* **2**, 4 (2009).
- Schmidt, C. & Hollmann, M. Molecular and functional characterization of *Xenopus laevis* *N*-methyl-D-aspartate receptors. *Mol. Cell. Neurosci.* **42**, 116–127 (2009).
- Karakas, E., Simorowski, N. & Furukawa, H. Structure of the zinc-bound amino-terminal domain of the NMDA receptor NR2B subunit. *EMBO J.* **28**, 3910–3920 (2009).
- Kumar, J., Schuck, P., Jin, R. & Mayer, M. L. The N-terminal domain of GluR6-subtype glutamate receptor ion channels. *Nature Struct. Mol. Biol.* **16**, 631–638 (2009).
- Jin, R. *et al.* Crystal structure and association behaviour of the GluR2 amino-terminal domain. *EMBO J.* **28**, 1812–1823 (2009).
- Clayton, A. *et al.* Crystal structure of the GluR2 amino-terminal domain provides insights into the architecture and assembly of ionotropic glutamate receptors. *J. Mol. Biol.* **392**, 1125–1132 (2009).
- Sobolevsky, A. I., Rosconi, M. P. & Gouaux, E. X-ray structure, symmetry and mechanism of an AMPA-subtype glutamate receptor. *Nature* **462**, 745–756 (2009).
- Gielen, M., Siegler Retchless, B., Mony, L., Johnson, J. W. & Paoletti, P. Mechanism of differential control of NMDA receptor activity by NR2 subunits. *Nature* **459**, 703–707 (2009).
- Rachline, J., Perin-Dureau, F., Le Goff, A., Neyton, J. & Paoletti, P. The micromolar zinc-binding domain on the NMDA receptor subunit NR2B. *J. Neurosci.* **25**, 308–317 (2005).
- Malherbe, P. *et al.* Identification of critical residues in the amino terminal domain of the human NR2B subunit involved in the RO 25-6981 binding pocket. *J. Pharmacol. Exp. Ther.* **307**, 897–905 (2003).
- Masuko, T. *et al.* A regulatory domain (R1–R2) in the amino terminus of the *N*-methyl-D-aspartate receptor: effects of spermine, protons, and ifenprodil, and structural similarity to bacterial leucine/isoleucine/valine binding protein. *Mol. Pharmacol.* **55**, 957–969 (1999).
- Perin-Dureau, F., Rachline, J., Neyton, J. & Paoletti, P. Mapping the binding site of the neuroprotectant ifenprodil on NMDA receptors. *J. Neurosci.* **22**, 5955–5965 (2002).
- Lee, C. H. & Gouaux, E. Amino terminal domains of the NMDA receptor are organized as local heterodimers. *PLoS ONE* **6**, e19181 (2011).
- Adams, P. D. *et al.* PHENIX: building new software for automated crystallographic structure determination. *Acta Crystallogr. D* **58**, 1948–1954 (2002).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank the staff at X25 and X29 at the National Synchrotron Light Source for beamline support. M. Mayer is thanked for comments on this work. We also thank D. Raleigh for the use of analytical ultracentrifugation. GluN1 clones from *Xenopus laevis* were gifts from M. Hollmann and H. Cline. This work was supported by NIH MH085926, the Alzheimer's Association and a donation from the Fox family (to H.F.). H.F. was also funded by a scientist development grant from the American Heart Association. E.K. is supported by a NARSAD Lieber Young Investigator Award.

**Author Contributions** The project was initiated by E.K. and H.F. All of the experiments were designed by E.K. and H.F. Crystallographic studies, isothermal calorimetry and analytical ultracentrifugation were carried out by E.K. Electrophysiology and crosslinking experiments were conducted by H.F. Technical support was given by N.S. The manuscript was written by H.F. and E.K.

**Author Information** Structural coordinates are deposited in the Protein Data Bank with accession codes 3QEK for GluN1b-ATD, 3QEL for GluN1b–GluN2B-ATD in complex with ifenprodil and 3QEM for GluN1b–GluN2B-ATD in complex with Ro 25-6981. Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at [www.nature.com/nature](http://www.nature.com/nature). Correspondence and requests for materials should be addressed to H.F. ([furukawa@cshl.edu](mailto:furukawa@cshl.edu)).

## METHODS

**Expression, purification and crystallization of GluN1b and GluN2B ATDs.**

The *Xenopus laevis* GluN1b ATD (Met1 to Glu408), containing Cys22Ser, Asn61Gln and Asn371Gln mutations, was C-terminally fused to a thrombin cleavage site followed by an octa-histidine tag. The *Xenopus laevis* GluN1b ATD and rat GluN2B ATD<sup>10</sup> constructs were individually expressed or co-expressed using the High Five (*Trichoplusia ni*) baculovirus system (DH10multibac)<sup>22</sup>. Purification was performed using a similar method to that described previously<sup>10</sup> except that the proteins were de-glycosylated by endoglycosidase F1 after purification by metal-chelate chromatography, and 1  $\mu$ M ifenprodil or 1  $\mu$ M Ro 25-6981 was included in the running buffer of size-exclusion chromatography (Superdex200) for isolation of the GluN1b–GluN2B ATD complex. The proteins used for isothermal titration calorimetry and sedimentation experiments were purified without the endoF1 de-glycosylation step and in the absence of ifenprodil or Ro 25-6981. Se-Met-incorporated GluN1b ATD proteins were expressed using methionine-free media (ESF921) supplemented with DL-Se-Met (Sigma) at 100 mg l<sup>-1</sup> (ref. 10). The GluN1b ATD and GluN1b–GluN2B ATDs were crystallized by hanging-drop vapour diffusion at 17 °C by mixing the protein (8 mg ml<sup>-1</sup>) at a 1:1 ratio with a reservoir solution containing 20% PEG3350, 150 mM KNO<sub>3</sub> and 50 mM HEPES-NaOH (pH 7.0) for GluN1b ATD, or at a 2:1 ratio with a solution containing 3.0–3.5 M sodium formate and 0.1 M HEPES (pH 7.5) for the GluN1b/GluN2B ATDs.

**Data collection and structural analysis.** Crystals were cryoprotected in buffers containing 20% PEG3350, 150 mM KNO<sub>3</sub>, 50 mM HEPES-NaOH (pH 7.0) and 20% glycerol for GluN1b ATD, or 5 M sodium formate and 0.1 M HEPES-NaOH (pH 7.5) for GluN1b–GluN2B ATDs. X-ray diffraction data were collected at the X25 and X29 beamlines at the National Synchrotron Light Source and processed using HKL2000 (ref. 23). Single anomalous diffraction data for the Se-Met-incorporated GluN1b ATD crystals were collected at the peak wavelength (0.9788 Å) and used for phasing by the program SHARP<sup>24</sup>. The initial model was built using flex-wARP<sup>25</sup>. The crystal structure of GluN1b–GluN2B ATD was solved by molecular replacement using the coordinates of GluN1b ATD and GluN2B ATD<sup>10</sup> (PDB code: 3JPW) with the program PHASER<sup>26</sup>. The models were built using COOT<sup>27</sup> and structural refinement was performed using the program PHENIX<sup>21</sup>.

**Isothermal titration calorimetry.** Proteins were dialysed overnight before the experiment against a buffer containing 150 mM NaCl, 20 mM Tris-HCl (pH 7.4) and 10% glycerol. Isothermal titration calorimetry measurements were performed using VP-ITC (MicroCal) by successive injections at 27 °C of 5  $\mu$ l of 0.15 mM ifenprodil to 0.01 mM GluN1b ATD, 10  $\mu$ l of 0.25 mM ifenprodil to 0.007 mM GluN2B ATD, 5  $\mu$ l of 0.15 mM ifenprodil to 0.01 mM GluN1b–GluN2B ATD complex and 5  $\mu$ l of 0.05 mM Ro 25-6981 to 0.007 mM GluN1b–GluN2B ATD complex. Data analysis was done using the software Origin 7.0 (Origin Labs).

**Analytical ultracentrifugation.** Sedimentation velocity and equilibrium experiments were performed using a Beckman Coulter Optima XL-I analytical ultracentrifuge. Proteins were dialysed against a buffer containing 150 mM NaCl and 20 mM Tris (pH 7.4), with or without 10  $\mu$ M ifenprodil. Sedimentation velocity experiments were performed by centrifuging protein samples loaded on 2-sector centrepieces at 42,000 r.p.m. at 20 °C. Concentration gradients were measured using interference optics or absorbance optics at a wavelength of 280 nm or 230 nm depending on the protein concentrations loaded (0.01, 0.05, 0.1, 0.5 and 1.2 mg ml<sup>-1</sup> for GluN1b–GluN2B ATD in the presence and absence of ifenprodil; 0.1 and 1.2 mg ml<sup>-1</sup> for GluN1b ATD and 0.1, 0.5 and 5 mg ml<sup>-1</sup> for GluN2B ATD). Data were analysed using the continuous c(s) and c(M) distribution models

implemented in Sedfit<sup>28</sup>. The weighted-average sedimentation coefficient ( $S_w$ ) was determined from the peak integration of c(s).

Sedimentation equilibrium experiments were performed using a 6-channel centrepiece loaded with 100- $\mu$ l protein samples at protein concentrations of 0.05, 0.1 and 0.3 mg ml<sup>-1</sup> in the presence or absence of 10  $\mu$ M ifenprodil. The samples were centrifuged sequentially at 9,000, 13,000 and 18,000 r.p.m. and allowed to reach equilibrium at each speed. Absorbance measurements were performed at wavelengths 230, 250 and 280 nm to obtain measurements at low and high protein concentrations. Global analysis of the data for multiple protein concentrations and rotor speeds was performed using single-species and A + B  $\leftrightarrow$  AB models implemented in Heteroanalysis v1.1.44 (University of Connecticut).

**Electrophysiology.** Recombinant GluN1/GluN2B NMDA receptors were expressed by co-injecting 0.1–0.5 ng of wild-type or mutant rat GluN1 and GluN2B cRNAs into defolliculated *Xenopus laevis* oocytes. The two-electrode voltage-clamp recordings were performed using agarose-tipped microelectrodes (0.4–1.0 M $\Omega$ ) filled with 3 M KCl at a holding potential of –40 mV. The bath solution contained 5 mM HEPES, 100 mM NaCl, 0.3 mM BaCl<sub>2</sub> and 10 mM Tricine at pH 7.4 (adjusted with KOH). Currents were evoked by the application of glycine and L-glutamate at 100  $\mu$ M each. Inhibition by ifenprodil was monitored in the presence of agonists and various concentrations of ifenprodil. For redox experiments, the oocytes were preincubated in the bath solution supplemented with 2 mM DTT for 3 min before recording in the continuous presence of 2 mM DTT. Data were acquired and analysed by the program Pulse (HEKA).

**Cysteine crosslinking and western blot.** Single point mutations were incorporated into the genes encoding full-length rat GluN1-4b and GluN2B in the pCI vector (Promega). Human embryonic kidney 293 cells were transfected by Eugene HD (Roche) with a mixture of 0.5  $\mu$ g of the GluN1-4b plasmid and 1  $\mu$ g of the GluN2B plasmid. Cells were harvested 24–48 h after transfection and resuspended in a buffer containing 20 mM Tris-HCl (pH 7.4), 150 mM NaCl, 1% dodecyl-maltoside and a protease-inhibitor cocktail (Roche), as previously described<sup>29</sup>. After centrifugation at 150,000g, the supernatant was subjected to SDS–polyacrylamide gel electrophoresis (4–15%) in the presence or absence of 100 mM DTT. The proteins were transferred to Hybond-ECL nitrocellulose membranes (GE Healthcare). The membranes were blocked with TBST (20 mM Tris-HCl (pH 7.4), 150 mM NaCl and 0.1% Tween-20) containing 10% milk, then incubated with mouse monoclonal antibodies against GluN1 (MAB 1586, Millipore) or GluN2B (Invitrogen), followed by HRP-conjugated anti-mouse antibodies (GE Healthcare). Protein bands were detected by ECL detection kit (GE Healthcare).

22. Fitzgerald, D. J. *et al.* Protein complex expression by using multigene baculoviral vectors. *Nature Methods* **3**, 1021–1032 (2006).
23. Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326 (1997).
24. de La Fortelle, E. & Bricogne, G. Maximum-likelihood heavy-atom parameter refinement for multiple isomorphous replacement and multiwavelength anomalous diffraction methods. *Methods Enzymol.* **276**, 472–494 (1997).
25. Cohen, S. X. *et al.* Towards complete validated models in the next generation of ARP/wARP. *Acta Crystallogr. D* **60**, 2222–2229 (2004).
26. McCoy, A. J. *et al.* Phaser crystallographic software. *J. Appl. Cryst.* **40**, 658–674 (2007).
27. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **60**, 2126–2132 (2004).
28. Schuck, P. Size-distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and Lamm equation modeling. *Biophys. J.* **78**, 1606–1619 (2000).
29. Furukawa, H., Singh, S. K., Mancusso, R. & Gouaux, E. Subunit arrangement and function in NMDA receptors. *Nature* **438**, 185–192 (2005).

# Coseismic and postseismic slip of the 2011 magnitude-9 Tohoku-Oki earthquake

Shinzauro Ozawa<sup>1</sup>, Takuya Nishimura<sup>1</sup>, Hisashi Suito<sup>1</sup>, Tomokazu Kobayashi<sup>1</sup>, Mikio Tobita<sup>1</sup> & Tetsuro Imakiire<sup>1</sup>

Most large earthquakes occur along an oceanic trench, where an oceanic plate subducts beneath a continental plate. Massive earthquakes with a moment magnitude,  $M_w$ , of nine have been known to occur in only a few areas, including Chile, Alaska, Kamchatka and Sumatra. No historical records exist of a  $M_w = 9$  earthquake along the Japan trench, where the Pacific plate subducts beneath the Okhotsk plate, with the possible exception of the AD 869 Jogan earthquake<sup>1</sup>, the magnitude of which has not been well constrained. However, the strain accumulation rate estimated there from recent geodetic observations is much higher than the average strain rate released in previous interplate earthquakes<sup>2–6</sup>. This finding raises the question of how such areas release the accumulated strain. A megathrust earthquake with  $M_w = 9.0$  (hereafter referred to as the Tohoku-Oki earthquake) occurred on 11 March 2011, rupturing the plate boundary off the Pacific coast of northeastern Japan. Here we report the distributions of the coseismic slip and postseismic slip as determined from ground displacement detected using a network based on the Global Positioning System. The coseismic slip area extends approximately 400 km along the Japan trench, matching the area of the pre-seismic locked zone<sup>4</sup>. The afterslip has begun to overlap the coseismic slip area and extends into the surrounding region. In particular, the afterslip area reached a depth of approximately 100 km, with  $M_w = 8.3$ , on 25 March 2011. Because the Tohoku-Oki earthquake released the strain accumulated for several hundred years, the paradox of the strain budget imbalance may be partly resolved. This earthquake reminds us of the potential for  $M_w \approx 9$  earthquakes to occur along other trench systems, even if no past evidence of such events exists. Therefore, it is imperative that strain accumulation be monitored using a space geodetic technique to assess earthquake potential.

Northeastern Japan has been struck by many M7-class ( $M_w = 7$ ) interplate earthquakes along the Japan trench, where the Pacific plate subducts beneath the Okhotsk plate at a rate of 73–78 mm yr<sup>-1</sup> (refs 7, 8; Fig. 1a). However, no interplate earthquake along the Japan trench with a surface-wave magnitude,  $M$ , of more than 7.5 has been instrumentally recorded since 1923, except along the northernmost part of the trench, where there have been  $M = 7.9$  and  $M = 7.6$  earthquakes (Fig. 1b). There are no historical records of any  $M_w > 8.5$  earthquakes occurring along the Japan trench since the seventeenth century. Therefore, the massive ( $M_w = 9$ ) Tohoku-Oki earthquake was not widely anticipated despite the geological evidence for recurrent devastating tsunamis in the past, in particular in AD 869<sup>1</sup>, and the rapid accumulation of elastic strain along the trench.

On the basis of earthquake catalogues covering several decades<sup>9</sup>, the seismic coupling coefficient, that is, the ratio between the rate of slip released in an interplate earthquake and the rate of relative plate motion, has been estimated to be 10–20% along the Japan trench<sup>2,3</sup>. However, ground displacement data acquired using a continuous Global Positioning System (GPS) network established in 1994 suggest that there is strong interplate coupling along the Japan trench<sup>4–6</sup> (Fig. 1b). The strain accumulation rate estimated from contemporary deformation is considerably higher than the average strain rate

released in historical earthquakes. An episodic aseismic slip, including an afterslip, has been suggested as a possible mechanism for significant elastic strain release<sup>4,10</sup>.

In this Letter, first we describe the coseismic and postseismic deformations associated with the Tohoku-Oki earthquake, as detected by the GPS Earth Observation Network (GEONET) operated by the Geospatial Information Authority of Japan<sup>11</sup>, and estimate the coseismic slip distribution and the subsequent afterslip distribution on the plate boundary by geodetic inversion<sup>12</sup> of the ground displacement at selected GPS sites. Second, we discuss the relationship between the coseismic and the postseismic slip models, as well as their relation to pre-seismic coupling and strain budget imbalance.

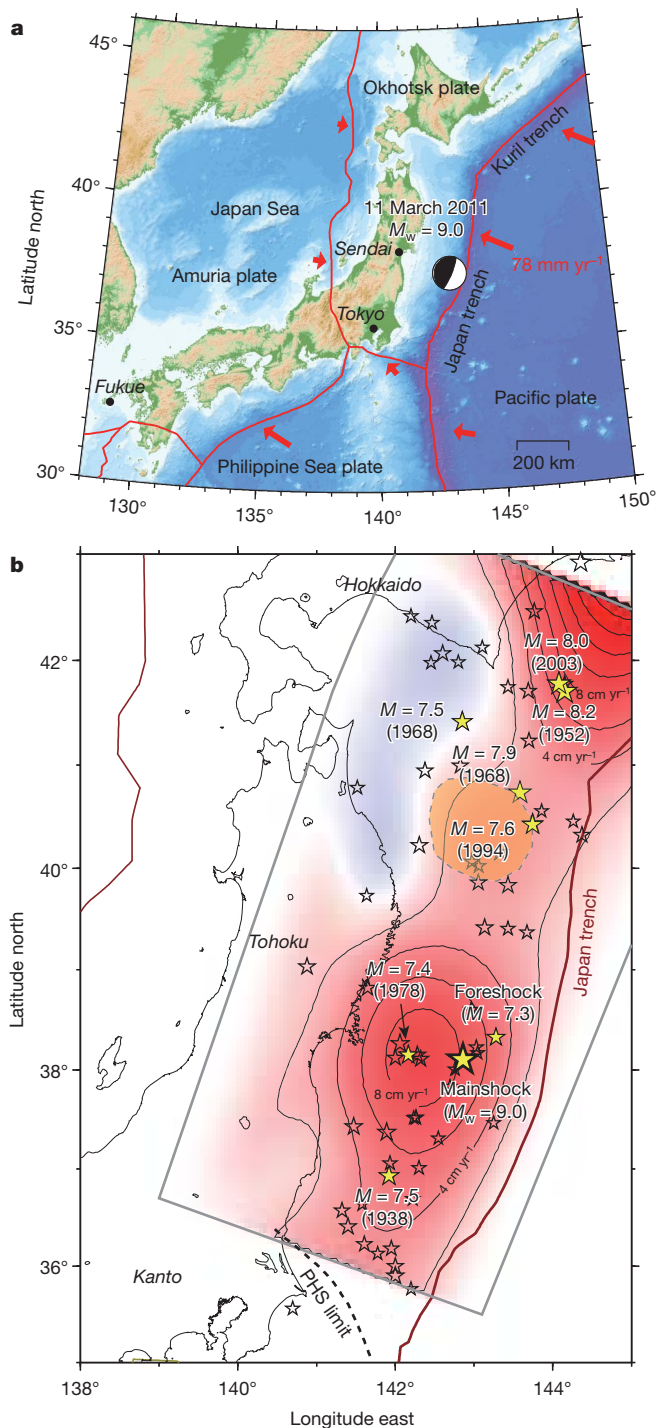
The observed coseismic displacements show eastward movements of up to 5.3 m and subsidence by up to 1.2 m along the coastal line of the Tohoku region, relative to the Fukue site (Figs 1 and 2a). These values are greater by one order of magnitude than those recorded at the time of previous M7–M8-class interplate earthquakes in the Tohoku region that have taken place since the establishment of GEONET. After the Tohoku-Oki earthquake, a large postseismic deformation occurred (Fig. 2b). Although the postseismic deformation resembles the coseismic field, the displacements seem to be more broadly distributed. In particular, the eastward displacement of the Pacific coastal area did not differ significantly from that of the western coastal area, whereas the eastward displacement on the Pacific coast was much larger than that of the western coastal area in the coseismic field. In addition, the Pacific coastal area near the source region was uplifted after the earthquake.

The slip distribution estimated on the basis of the coseismic displacements shows a large slip of up to 27 m near the epicentral area, extending approximately 400 km along the Japan trench at a depth of less than 60 km, which is the lower limit for the seismogenic zone along the subducting plate in this region<sup>13</sup> (Figs 1b and 2a). The estimated moment is  $3.43 \times 10^{22}$  N m, assuming a uniform rigidity of 40 GPa, equivalent to that of an  $M_w = 9.0$  earthquake. A uniform rigidity of 40 GPa is a rough average of 29, 41 and 50 GPa for the typical rigidities of upper crust, lower crust and upper mantle in northeastern Japan based on seismic data<sup>14</sup>. The moment of our geodetic model closely matches the moment magnitude of 9.1 inferred from seismic waveform analysis in ref. 15. The root mean squared deviation of this model is 0.011 m (Supplementary Figs 1 and 2), and the estimated slip is well beyond the  $1\sigma$  error (Supplementary Figs 3). Chequerboard and sensitivity tests show that spatial variations of slip over the coseismic area are resolved to the scale of a few tens of kilometres and that the principal pattern is a stable feature.

The estimated afterslip, which is based on the postseismic deformation, occurs in the coseismic slip area and adjacent to it, expanding to the north, the south and in the dipping direction (Figs 2b and 3). The afterslip area has two modal centres: northwest of the centre of the coseismic slip and east of the Kanto region. These centres reflect the large postseismic displacement along the Pacific coast, which, unlike the coseismic displacement, extends north and south (Fig. 2). The root mean squared deviation of this model is 0.007 m (Supplementary Figs 3 and 4). The estimated moment of the afterslip on 25 March is

<sup>1</sup>Geospatial Information Authority of Japan, Tsukuba, Ibaraki 305-0811, Japan.





**Figure 1 | Tectonic setting in and around the Tohoku-Oki earthquake.**

**a**, Plate configurations of the Japanese islands<sup>29</sup>. The focal mechanism of the Tohoku-Oki earthquake is taken from the Global Centroid-Moment-Tensor Project<sup>15</sup>. The red arrows indicate relative motion between the two plates at a plate boundary<sup>7,8</sup>. **b**, Coupling distribution before the earthquake and recent seismicity along the Japan trench. The colour shading and contours indicate the degree of interplate coupling between the subducting Pacific plate and the overriding Okhotsk plate, estimated from GPS data recorded between April 2000 and March 2001<sup>4</sup>. The degree of coupling is expressed as the backslip rate<sup>30</sup>, which is a slip deficit from the relative plate velocity. The stars mark the epicentres of large ( $M \geq 6.8$ ) earthquakes that have occurred since 1923. The epicentres of the mainshock, a foreshock and earthquakes with  $M \geq 7.4$  are marked by yellow stars and labelled with their magnitudes and/or times of occurrence. The orange area is the source area of the  $M = 7.6$  1994 earthquake<sup>20</sup>. The dashed line shows the northeastern limit of the subducted Philippine Sea plate<sup>21</sup> (PHS). The Okhotsk plate overrides the Pacific plate north of this limit and the Philippine Sea plate overrides the Pacific plate south of this limit. The grey rectangle represents a fault patch to estimate the backslip rate.

afterslip area, avoiding the large coseismic slip area (Fig. 3). This is consistent with the observation that in many cases the aftershocks and the large afterslip occur in an area where the coseismic slip is not large<sup>18,19</sup>. The seismic moments of thrust-type aftershocks sum to  $1.5 \times 10^{19}$  N m for the period of the postseismic deformation. This suggests that aftershocks contribute less than 1% of the moment of the afterslip model for two weeks. Although the estimated area of small afterslip overlaps the coseismic slip, we cannot rule out the possibility that this may be due to oversmoothing in the afterslip estimation, because a sensitivity test of the smoothing constraint shows that the afterslip area avoids the large coseismic slip area in undersmoothed models. The expansion in the dipping direction reaches 80–100 km in depth, which is the lower limit for the coupling of the plates in this region.

The propagation of the afterslip into an area deeper than the coseismic slip area was observed for the 1994 Sanriku-Haruka-Oki earthquake<sup>20</sup> ( $M = 7.6$ ) and the 2003 Tokachi-Oki earthquake<sup>18</sup> ( $M = 8.0$ ), suggesting a general dipping expansion along the Japan and Kuril trenches. Because the interplate coupling rate is near zero at depths of more than 100 km, we think that the afterslip area terminates at this limit.

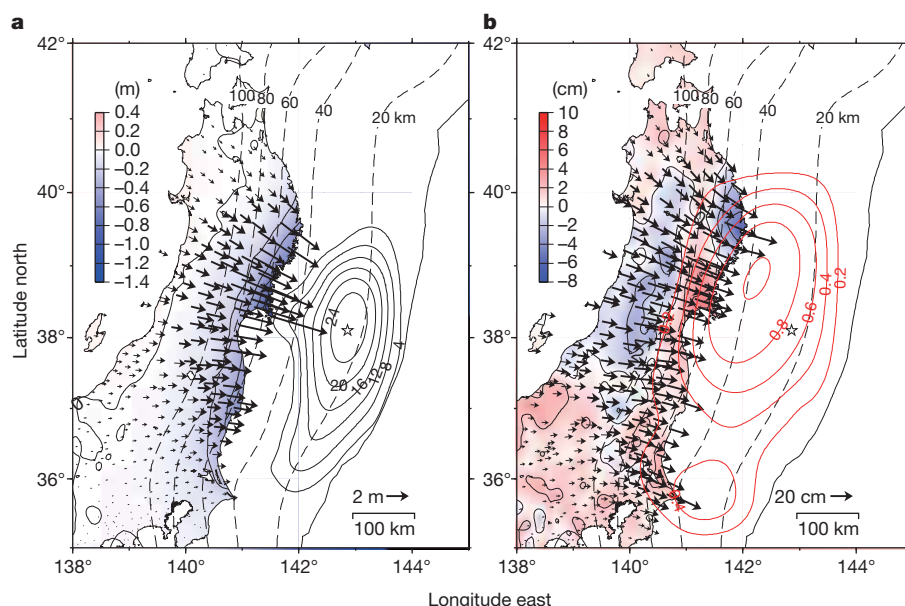
The northward expansion of the afterslip area approaches the zone ruptured in the 1968 ( $M = 7.9$ ) and 1994 ( $M = 7.6$ ) earthquakes, (Fig. 3). The afterslip area may terminate there because the source area of the 1994 earthquake is now strongly locked. The southward expansion has reached the Kanto region (Figs 1b, 2b and 3). The Philippine Sea plate overrides the Pacific plate in the Kanto region, south of the northeastern limit of the former plate, whereas the Okhotsk plate lies on the Pacific plate north of this limit<sup>21</sup> (Figs 1 and 3). There is a possibility that this change in the overriding plate stops the southward expansion of the afterslip at the limit of the Philippine Sea plate in the Kanto region. Our model estimates the afterslip distribution at intervals of 1 d, and the two modal centres of the afterslip area seem not to move significantly on this timescale whereas the slip magnitude increases rapidly.

The ruptured area of the Tohoku-Oki earthquake well matches the area estimated to have been strongly coupled before the earthquake<sup>4</sup> (Fig. 1b), although the centre of the coseismic slip area is shallower than that of the locked area. This finding indicates the extreme importance of the GPS observations in assessing the potential of a subduction earthquake occurring, as has been observed in other subduction zones<sup>22</sup>. A deeper part of the locked zone may release the remaining strain energy by further afterslip of the Tohoku-Oki earthquake.

The moment accumulation rate attributed to the subduction of the Pacific plate in an area from latitude  $36^\circ$  N to  $39.5^\circ$  N along the Japan trench is estimated to have been approximately  $1.6 \times 10^{20}$  N m  $\text{yr}^{-1}$  before the earthquake<sup>4</sup>. The repeated occurrence of  $M_w < 8$  interplate earthquakes contributes 10–20% of the plate motion, without taking the afterslip into account<sup>2,3</sup>. It remains unclear how much energy will be released by the afterslip of the Tohoku-Oki earthquake. In the case

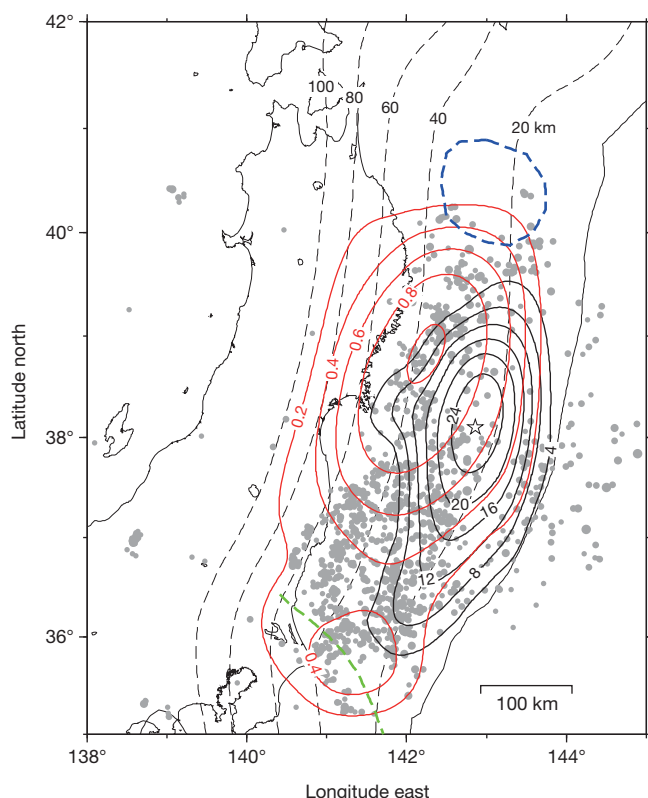
$3.35 \times 10^{21}$  N m, which is equivalent to that of an  $M_w = 8.3$  earthquake (Figs 2b and 3). This moment is approximately 10% of that of the mainshock. We assume that the postseismic deformation transients are due solely to afterslip, although they are affected by the viscoelastic relaxation of the asthenosphere and poroelastic rebound<sup>16</sup>. We estimate the magnitude of these effects by simple calculation. The viscoelastic relaxation model<sup>17</sup> is predicted to be within 1 cm of surface displacement for two weeks, with an asthenospheric viscosity of  $10^{19}$  Pa. Although poroelastic effects reach 20% of the observed post-seismic deformation at maximum, their horizontal and vertical patterns are quite different from the observations. Thus, we assume that these effects can be ignored as a first approximation in this case.

The area of large afterslip is located in the region peripheral to the coseismic slip zone. In addition, aftershocks seem to occur in the



**Figure 2 | Coseismic and postseismic displacements and estimated slip.** **a**, Coseismic displacements for 10–11 March 2011, relative to the Fukue site. The black arrows indicate the horizontal coseismic movements of the GPS sites. The colour shading indicates vertical displacement. The star marks the location of the earthquake epicentre. The dotted lines indicate the isodepth contours of

the plate boundary at 20-km intervals<sup>28</sup>. The solid contours show the coseismic slip distribution in metres. **b**, Postseismic displacements for 12–25 March 2011, relative to the Fukue site. The red contours show the afterslip distribution in metres. All other markings represent the same as in **a**.



**Figure 3 | Coseismic slip, postseismic slip and aftershocks.** Estimated coseismic slip (black contour, 4-m interval) and postseismic slip (red contour, 0.2-m interval) of the Tohoku-Oki earthquake for the same period as in Fig. 2. The green dashed line indicates the northeast limit of the Philippine Sea plate. The blue dashed line indicates the ruptured area of the  $M = 7.6$  1994 earthquake<sup>20</sup>. The grey circles show the epicentres of the aftershocks of the Tohoku-Oki earthquake for 11–25 March 2011. All other markings represent the same as in Fig. 2.

of the 1994 Sanriku-Haruka-Oki earthquake, the afterslip released moment equivalent to that of the mainshock over the course of 1 yr (ref. 10). In other cases, it has been reported that the postseismic slip releases a moment of up to  $\sim 30\%$  of that of the mainshock over several weeks after M8-class earthquakes<sup>23</sup>. The moment released by afterslip after the 2004 Sumatra earthquake is also estimated to have been 30% of the moment of the mainshock over 40 d (ref. 24). If we assume that the afterslip of the Tohoku-Oki earthquake eventually releases 30–100% of the amount of energy released by the mainshock, and extrapolate the interplate coupling from the geodetic observations, we find that it would take approximately 350–700 yr for energy equivalent to that of the earthquake to accumulate along the Japan trench.

Recent geological studies suggest that tsunamis similar to that which followed the Tohoku-Oki earthquake have repeatedly struck the Pacific coast of northeastern Japan, with a recurrence interval of approximately 800–1,100 yr (ref. 1), implying that megathrust earthquakes have also occurred repeatedly along the Japan trench. The massive Tohoku-Oki earthquake supports this hypothesis and may partly resolve the strain budget imbalance, although the roughly estimated recurrence interval is shorter than the tsunami return period.

The Pacific coastal area subsided by up to 1.2 m following the earthquake. Furthermore, subsidence of the Pacific coast at a rate of 5–10 mm yr<sup>-1</sup> over the past 100 yr has been estimated by tide gauges<sup>25</sup>, levelling and GPS data. Although geodetic observations indicate both coseismic and interseismic subsidence, a geomorphological study has shown that there was long-term upheaval along the Pacific coast in the late Quaternary period<sup>26</sup>. This discrepancy suggests the existence of another mechanism of episodic uplift, such as postseismic deformation<sup>27</sup>. In fact, the Pacific coastal area near the epicentre started to uplift after the Tohoku-Oki earthquake by an amount ranging from 1 to 4 cm, as observed over the course of two weeks (Fig. 2). It is difficult to predict the future temporal evolution of uplifting from such a short observation period. If the uplift lasts for a long time, the discrepancy between subsidence and upheaval will be resolved. To understand the uplift mechanism, it is important to continue geodetic monitoring.

## METHODS SUMMARY

Coseismic and postseismic displacements were based on GPS data collected for 6 h and were analysed with the BERNESE GPS software. We used the east–west,

north–south and up–down components, measured relative to the Fukue site, at approximately 400 selected GPS sites covering northeastern Japan (Fig. 1a). We used the Yabuki–Matsu'ura method<sup>12</sup> to estimate the slip distribution on the plate boundary, and used a fault patch that covers an area ~500 km in width and ~800 km in length to represent the plate boundary in this region<sup>28</sup>. The fault patch was represented by a parametric spline surface. Green's functions<sup>12</sup> that assume a homogeneous half-space were used. A detailed description of the GPS inversion approach, including resolution and sensitivity analysis, can be found in Methods and Supplementary Figs 5–9. The data set and the results of inversion are shown in Supplementary Tables 1–4.

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Received 31 March; accepted 25 May 2011.**

**Published online 15 June 2011.**

- Minoura, K., Imamura, F., Sugawara, D., Kono, Y. & Iwashita, T. The 869 Jogan tsunami deposit and recurrence interval of large-scale tsunami on the Pacific coast of northeast Japan. *J. Nat. Disaster Sci.* **23**, 83–88 (2001).
- Peterson, E. T. & Seno, T. Factors affecting seismic moment release rates in subduction zones. *J. Geophys. Res.* **89**, 10233–10248 (1984).
- Pacheco, J. F., Sykes, L. R. & Scholz, C. H. Nature of seismic coupling along simple plate boundaries of the subduction type. *J. Geophys. Res.* **98**, 14133–14159 (1993).
- Nishimura, T. *et al.* Temporal change of interplate coupling in northeastern Japan during 1995–2002 estimated from continuous GPS observations. *Geophys. J. Int.* **157**, 901–916 (2004).
- Hashimoto, C., Noda, A., Sagiya, T. & Matsu'ura, M. Interplate seismogenic zones along the Kuril–Japan trench from GPS data inversion. *Nature Geosci.* **2**, 141–144 (2009).
- Suwa, Y., Miura, S., Hasegawa, A., Sato, T. & Tachibana, K. Interplate coupling beneath NE Japan inferred from three-dimensional displacement field. *J. Geophys. Res.* **111**, B04402 (2006).
- Sella, G. F., Dixon, T. H. & Mao, A. REVEL: a model for recent plate velocities from space geodesy. *J. Geophys. Res.* **107**, 2081 (2002).
- Apel, E. V. *et al.* Independent active microplate tectonics of northeast Asia from GPS velocities and block modeling. *Geophys. Res. Lett.* **33**, L11303 (2006).
- Utsu, T. Chronological table of earthquakes in Japan with a moment magnitude larger than 6.0 and disastrous earthquakes from 1885 to 1980. *Bull. Earthq. Res. Inst.* **57**, 401–463 (1983).
- Heki, K., Miyazaki, S. & Tsuji, H. Silent fault slip following an interplate thrust earthquake at the Japan Trench. *Nature* **386**, 595–597 (1997).
- Sagiya, T., Miyazaki, S. & Tada, T. Continuous GPS array and present-day crustal deformation of Japan. *Pure Appl. Geophys.* **157**, 2303–2322 (2000).
- Yabuki, T. & Matsu'ura, M. Geodetic data inversion using a Bayesian information criterion for spatial distribution of fault slip. *Geophys. J. Int.* **109**, 363–375 (1992).
- Igarashi, T., Matsuzawa, T., Umino, N. & Hasegawa, A. Spatial distribution of focal mechanisms for and intraplate earthquakes associated with the subducting Pacific plate beneath the northeastern Japan arc: a triple-planed deep seismic zone. *J. Geophys. Res.* **106**, 2177–2191 (2001).
- Nakajima, J. & Matsuzawa, T. Hasegawa, A. & Zhao, D. Seismic imaging of arc magma and fluids under the central part of northeast Japan. *Tectonophysics* **341**, 1–17 (2001).
- Global CMT Web Page. *Global Centroid Moment Tensor Project* (<http://www.globalcmt.org/>) (accessed 30 March 2011).
- Shearer, P. & Burgmann, R. Lessons learned from the 2004 Sumatra–Andaman megathrust rupture. *Annu. Rev. Earth Planet. Sci.* **38**, 103–131 (2010).
- Pollitz, F. Gravitational viscoelastic postseismic relaxation on a layered spherical earth. *J. Geophys. Res.* **102**, 17921–17941 (1997).
- Ozawa, S., Kaidzu, M., Murakami, M., Imakiire, T. & Hatanaka, Y. Coseismic and postseismic crustal deformation after the  $M_w$  8 Tokachi-oki earthquake in Japan. *Earth Planets Space* **56**, 675–680 (2004).
- Hsu, Y. *et al.* Frictional afterslip following the 2005 Nias–Simeulue earthquake, Sumatra. *Science* **312**, 1921–1926 (2006).
- Nishimura, T. *et al.* Distribution of seismic coupling on the subducting plate boundary in northeastern Japan inferred from GPS observations. *Tectonophysics* **323**, 217–238 (2000).
- Uchida, N., Matsuzawa, T., Nakajima, J. & Hasegawa, A. Subduction of a wedge-shaped Philippine Sea plate beneath Kanto, central Japan, estimated from converted waves and small repeating earthquakes. *J. Geophys. Res.* **115**, B07309 (2010).
- Moreno, M., Rosenau, M. & Oncken, O. 2010 Maule earthquake slip correlates with pre-seismic locking of Andean subduction zone. *Nature* **467**, 198–202 (2010).
- Melbourne, T. I., Webb, F. H., Stock, J. M. & Reigbar, C. Rapid postseismic transients in subduction zones from continuous GPS. *J. Geophys. Res.* **107**, 2241 (2002).
- Chlieh, M. *et al.* Coseismic slip and afterslip of the great  $M_w$  9.15 Sumatra–Andaman earthquake of 2004. *Bull. Seismol. Soc. Am.* **97**, S152–S173 (2007).
- Kato, T. Secular and earthquake-related vertical crustal movements in Japan as deduced from tidal records (1951–1981). *Tectonophysics* **97**, 183–200 (1983).
- Matsu'ura, T., Furusawa, A. & Saomoto, H. Long-term and short-term vertical velocity profiles across the forearc in the NE Japan subduction zone. *Quat. Res.* **71**, 227–238 (2009).
- Suito, H. & Freymueller, J. T. A viscoelastic and afterslip postseismic deformation model for the 1964 Alaska earthquake. *J. Geophys. Res.* **114**, B11404 (2009).
- Nakajima, J. & Hasegawa, A. Anomalous low-velocity zone and linear alignment of seismicity along it in the subducted Pacific slab beneath Kanto, Japan: reactivation of subducted fracture zone? *Geophys. Res. Lett.* **33**, L16309 (2006).
- Bird, P. An updated digital model of plate boundaries. *Geochem. Geophys. Geosyst.* **4**, 1027 (2003).
- Savage, J. C. A dislocation model of strain accumulation and release at a subduction zone. *J. Geophys. Res.* **88**, 4984–4996 (1983).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We are grateful to the Japan Meteorological Agency for providing us with the hypocentre data. We also thank M. Murakami, H. Munekane and Y. Hatanaka for their comments.

**Author Contributions** S.O., T.N. and H.S. participated in the construction of the model and wrote the manuscript. M.T., T.J. and T.K. participated in the discussion of the results and reviewed the manuscript.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at [www.nature.com/nature](http://www.nature.com/nature). Correspondence and requests for materials should be addressed to S.O. ([ozawa@gsi.go.jp](mailto:ozawa@gsi.go.jp)).



## METHODS

The objective of our study was to determine the distribution of the coseismic and postseismic slip of the Tohoku-Oki earthquake, Japan. The analysis was based on the modelling of the coseismic and postseismic deformation by a GPS network in Japan. A solution was obtained by inverting a set of 377 coseismic slip GPS vectors and 357 postseismic slip GPS vectors along the Japan trench.

**GPS data.** The GPS data used in this study were derived by the most rapid of the three strategies in the baseline analysis of the GEONET routine<sup>31,32</sup>. In the strategy for this solution, 6-h data are processed to estimate the static coordinates using the software BERNES 5.0 and the ultrarapid ephemerides of the International Global Navigation Satellite Systems Service (IGS) every 3 h. Tropospheric zenith delays and gradients for two and, respectively, a single time segment were estimated using the Niell mapping function<sup>33</sup> in each session. The elevation cut-off angle for the GPS satellites was 15°, and absolute phase centre corrections to the GPS antennas were applied for each monument type of the GEONET stations. Thus, the GPS carrier-phase ambiguities were resolved. The site coordinates were compared with the 2005 International Terrestrial Reference Frame<sup>34</sup> (ITRF2005) by fixing a fiducial station (station 92110, close to the Tsukuba IGS station) as the a-priori value using ITRF2005. Because of the large coseismic displacement of the Tohoku-Oki earthquake, the fiducial station moved eastwards by ~0.5 m. In the baseline analysis after the earthquake, the a-priori coordinate of the fiducial station was corrected by adding the coseismic offset. The coseismic and postseismic displacements were calculated from the relative site coordinates with respect to the reference site (Fukue; station code 950462), which is ~1,400 km away from the earthquake epicentres. The repeatabilities of the relative coordinates were 4.0, 3.6 and 15.1 mm for the east–west, north–south and up–down components, respectively, which are averages of standard deviations for the 412 sites used during 1–10 February 2011.

**Model strategy.** Coseismic offsets were estimated by subtracting the average coordinates for the period between 21:00 on 9 March and 09:00 on 11 March from the coordinates at 18:00 on 11 March 2011 (Japan Standard Time). According to the repeatabilities of the coordinates, the errors in the coseismic offsets were approximately 6 and 20 mm in the horizontal and vertical components, respectively. Postseismic data on 25 March were estimated by subtracting coordinates at 18:00 on 11 March from those at 18:00 on 25 March. We used the same error estimates as those for the coseismic deformation. We used 377 and 357 GPS sites in inversion for coseismic slip and postseismic slip, respectively.

We created a parametric spline surface to represent the surface of the subducting Pacific plate in the Tohoku region<sup>28,35</sup>. The parametric spline surface consisted of 15 knots in the dipping direction and 25 knots along the Japan trench. An adopted spline surface covers approximately 500 km in width, 800 km in length and 200 km in depth.

We used an inversion method<sup>12</sup> with minor modifications. In our inversion method, east–west and north–south slip components are represented by a parametric spline surface, as is the case for a fault patch. The vertical slip component is estimated using the formula in ref. 12. Green's functions are calculated using the formulation of ref. 12, which assumes a homogeneous isotropic half-space. We included a roughness matrix<sup>12</sup> as prior information, imposing the condition  $\lambda_1 \mathbf{M} \mathbf{u} = \mathbf{0}$  on the inversion equation, where  $\lambda_1$  is a hyperparameter of roughness,  $\mathbf{M}$  is the roughness matrix<sup>12</sup> and  $\mathbf{u}$  represents slip. We also adopted the prior information that the slip direction is parallel to the motion of the subducting Pacific plate, by imposing the condition  $\lambda_2 (\mathbf{u}_1 - \mathbf{u}_2) = \mathbf{0}$  on the inversion equation, where  $\mathbf{u}_1$  and  $\mathbf{u}_2$  are slip vectors angled at 45° relative to the direction of plate motion and  $\lambda_2$  is a hyperparameter. These roughness and slip rake constraints result in a stable solution. By including the roughness and slip directions, we were able to estimate the two hyperparameters by minimizing the Akaike Bayesian information criterion<sup>12,36</sup>. Minimization was done using Powell's method<sup>37</sup>. We set to zero the displacements at the edge of the fault surface and the displacements east of the Japan trench, as a boundary condition.

**Resolution and sensitivity of the inversion.** We tested the resolution power by attempting to recover a given coseismic slip distribution, which is often called

a chequerboard resolution test. We discretized the fault plane into regular chequerboard patterns with patches assigned either 0 or 5 m eastward interplate slip for the coseismic case (Supplementary Fig. 5a). A forward model introduces the chequerboard pattern as a fault slip condition and then simulates the displacements at the GPS sites. Gaussian random noises corresponding to GPS measurement errors were then added to obtain a set of synthetic data, which we subsequently inverted. The initial chequerboard pattern used in the test had an approximate size of 100 km × 100 km, which dimensions are similar to the wavelength of the structures we want to resolve, that is, the coseismic slip. The preferred models in the coseismic case recovered the main features of the chequerboard in most parts of the model region (Supplementary Fig. 5b). Resolution was generally good in the down-dip and onshore regions but was relatively poor near the trench. A chequerboard pattern was almost reproduced in the western part of the large coseismic slip area. This indicates that our inversion method was able to recover the main features of the coseismic slip distribution in the Tohoku-Oki earthquake.

By using a set of roughness coefficients, ranging from the oversmoothed (Supplementary Fig. 6a–c) to the undersmoothed (Supplementary Fig. 6e–f), we tested the sensitivity of the slip distribution to the roughness coefficient,  $\lambda_1$ . Because  $\lambda_2$ , which constrains the slip direction, does not significantly affect the slip distribution, we show the sensitivity only for  $\lambda_1$ . Supplementary Fig. 6d shows the optimal slip model, as determined using the minimum Akaike Bayesian information criterion. This sensitivity test suggests that the estimated coseismic slip distribution characterized by the area of large slip east of Sendai and the ~200–400-km-long slip area along the Japan trench is a robust feature that does not depend on the chosen roughness of the slip distribution.

We conducted similar tests for the postseismic slip. Supplementary Fig. 7a shows the initial chequerboard pattern, in which we assigned 0.4 and 0 m eastward interplate slip to the pink and blue areas, respectively. The resulting model shows good recovery of the chequerboard pattern, especially in the large afterslip area of Figs 2 and 3, but it suggests relatively poor recovery in the offshore area, which is 100 km away from land. Supplementary Fig. 8d depicts the optimal slip model. This sensitivity test indicates that the centre of the afterslip area is located in the down-dip area of the coseismic slip region, which is independent of the roughness constraint. However, the undersmoothed model (Supplementary Fig. 8e, f) shows that the afterslip concentrates along the down-dip edge of the large coseismic area.

We also tested a situation in which the afterslip does not overlap the coseismic slip area, and checked whether the afterslip distribution assigned in this case was reproduced by our inversion. The assumed condition is the same as the chequerboard test for the postseismic case, except for a given slip distribution. The results show that the assigned slip is well recovered by inversion, although the area of small slip is extended into the coseismic area owing to the smoothness constraint (Supplementary Fig. 9). Thus, we conclude that the slip centre shifts to a deeper portion of the coseismic rupture area in the postseismic period. However, owing to the limited resolving power of the geodetic data, it is not clear whether the afterslip area overlaps the coseismic slip area.

31. Hatanaka, Y. *et al.* Improvement of the analysis strategy of GEONET. *Bull. Geogr. Surv. Inst.* **49**, 11–37 (2003).
32. Nakagawa, H. Development and validation of GEONET new analysis strategy (version 4) [in Japanese]. *J. Geogr. Surv. Inst.* **118**, 1–8 (2009).
33. Niell, A. E. Global mapping functions for the atmosphere delay at radio wavelengths. *J. Geophys. Res.* **101**, 3227–3246 (1996).
34. Altamimi, Z., Collilieux, X., Legrand, J., Garayt, B. & Boucher, C. ITRF2005: a new release of the International Terrestrial Reference Frame based on time series of station positions and Earth orientation parameters. *J. Geophys. Res.* **112**, B09401 (2007).
35. Ozawa, S., Murakami, M. & Tada, T. Time-dependent inversion study of the slow thrust event in the Nankai trough subduction zone, southwestern Japan. *J. Geophys. Res.* **106**, 787–802 (2001).
36. Akaike, H. A new look at the statistical model identification. *IEEE Trans. Automat. Contr.* **AC-19**, 716–723 (1974).
37. Press, H. W., Teukolsky, S. A., Vetterling, W. T. & Flannery, B. P. *Numerical Recipes in Fortran* 409–411 (Cambridge Univ. Press, 1992).

## EARTHQUAKES

# The lessons of Tohoku–Oki

An exceptional data set documents surface deformation before, during and after the earthquake that struck northeastern Japan in March 2011. But models for assessing seismic and tsunami hazard remain inadequate.

JEAN-PHILIPPE AVOUAC

Earthquake science has entered a new era with the development of space-based technologies to measure surface deformation at the boundaries of tectonic plates and large faults. Japan has been at the forefront in implementing these technologies, in particular with the deployment some 15 years ago of a network of continuously recording Global Positioning System (GPS) stations known as GeoNet. Papers analysing the data associated with the devastating Tohoku–Oki earthquake of 11 March 2011 are now appearing. The latest of these, by Ozawa *et al.*<sup>1</sup>, is published online in *Nature* today.

With a moment magnitude ( $M_w$ ) of 9.0, the Tohoku–Oki earthquake ranks among the largest ever recorded. The data collected at the GeoNet stations<sup>1</sup> indicate that it resulted from the sudden slip of a remarkably compact area (400 kilometres long by 200 kilometres wide) of the plate interface where the Pacific plate slides beneath the Okhotsk plate, on which northern Japan lies. The rupture area (Fig. 1) lies off the coast-line of Honshu, Japan's biggest island, and extends east nearly all the way to the Japan trench — hence the particularly devastating tsunami produced by the earthquake.

Other new papers<sup>2,3</sup> provide further information. A combination<sup>2</sup> of GPS measurements and underwater acoustic sounding shows that the sea bottom in the epicentral area moved seaward by as much as 24 metres and was uplifted by about 3 metres. Therefore, slip along the plate interface at depth must have exceeded the 27-metre peak slip inferred from the GeoNet data<sup>1</sup>; it may have been even more than 50 metres, as suggested from the joint modelling<sup>3</sup> of the GeoNet data and sea-bottom pressure records of the tsunami waves. For comparison, this is about twice the peak slip determined for the giant earthquakes in Sumatra in 2004 ( $M_w$  9.4) and in Chile in 2010 ( $M_w$  9.0) — and larger than that estimated for the biggest earthquake ever recorded<sup>4</sup>, the

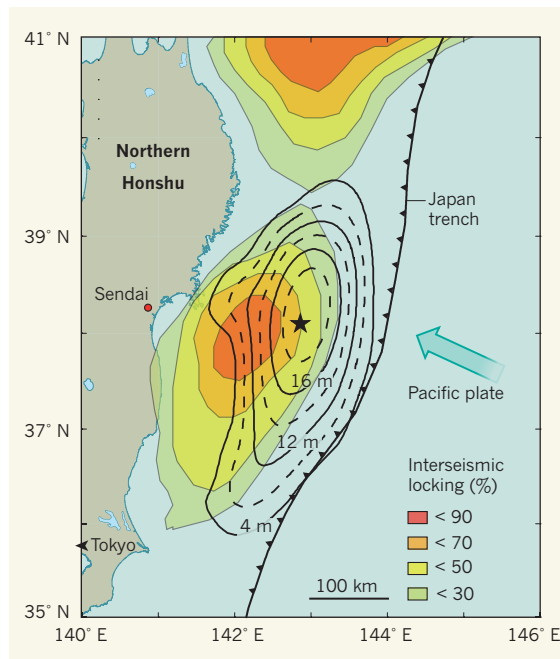
$M_w$ -9.5 event of 1960, which ruptured more than 1,000 km of the plate boundary off the coast of southern Chile.

Over the 15 years preceding the 2011 event, the GeoNet data<sup>5</sup> had revealed the slow accumulation of strain across Honshu, with the Pacific plate squeezing and dragging down the eastern edge of Honshu. We know, however, that the coast of Honshu is being uplifted in the long term, so a significant fraction of that 'interseismic' strain — strain accumulating between earthquakes — must be compensated by sudden episodes of uplift. The current model holds that interseismic strain on the

upper plate is purely elastic, and is 'recovered' during seismic rupture of the plate interface, so that in the long run the upper plate does not deform<sup>6</sup>. This assumption provides a rationale to relate slip on the plate interface to interseismic strain on the upper plate. Where the plate interface is creeping, strain on the upper plate is negligible; but where the plate interface is locked, the upper plate is compressed and dragged down, building up elastic strain until it is released when the locked patches slip.

Several earlier studies adopted this assumption<sup>5,7,8</sup> and found that the measured strain across Honshu required a large, locked patch off the coast of Sendai (Fig. 1). The rupture area of the Tohoku–Oki earthquake coincides quite well with that area<sup>3</sup>. A noteworthy discrepancy, however, is that the rupture reached closer to the Japan trench, where interseismic models suggested there was little locking. The particularly extensive, shallow slip seen in the Tohoku–Oki event could be due either to high pre-seismic stress left over from previous ruptures of the plate interface that failed to reach the trench, or, as seismological investigations suggest<sup>9</sup>, to specific properties of the plate interface. In any case, the observed slip requires the shallow plate interface to have remained locked, at least partially, in the period before the earthquake.

The published interseismic models<sup>5,7,8</sup> indicated little locking at shallow depth, essentially as a result of built-in methodological assumptions: the shallow portion of the plate interface is actually not well constrained if only onshore data are used<sup>1</sup>. These models may have been misleadingly interpreted as discounting the possibility of extensive shallow slip, which in fact occurred. Therefore, in the absence of direct constraints from sea-bottom geodesy, it may be preferable for models to assume maximum locking of the shallow zone of the plate interface. In fact, the interseismic data do not exclude a locked region off the coast of Sendai extending all the way to the shallow plate-boundary zone at the trench. Such an assumption raises questions for



**Figure 1 | Location of the Tohoku–Oki earthquake.** The earthquake, with its epicentre marked by a star, ruptured the plate interface along which the Pacific plate slides beneath northern Honshu at a rate of 8 centimetres per year. Ozawa and colleagues' analysis<sup>1</sup> shows that the rupture area and distribution of slip (represented by the black contour lines) roughly coincide with a patch of the plate interface that had remained locked over the preceding decades<sup>8</sup> (coloured area east of Sendai). The earthquake source was extremely compact and produced very large slip at relatively shallow depth (less than 20 km depth), hence the devastating tsunami. The other well-locked patch in the north coincides with rupture areas of large historical earthquakes (in particular, the  $M_w$ -8.5 Sanriku earthquake of 1896 and the  $M_w$ -8.2 Tokachi–Oki earthquake of 1968).

assessing the frequency of Tohoku-Oki-like earthquakes.

The estimated slip along the plate boundary off the coast of northern Honshu — due to earthquakes over the past few centuries — falls well short of balancing the slip deficit that should have accumulated over that period owing to interseismic locking. So it might seem that a large earthquake there was overdue. Indeed, according to the published interseismic models, interseismic strain builds up really fast on that boundary: it should take only a few centuries to accumulate enough strain to generate an  $M_w$ -9.0 earthquake. Such large events should recur even more often if locking of the shallow portion of the plate interface is assumed in the modelling of interseismic strain.

By contrast, on the basis of historical and palaeo-tsunami records<sup>10</sup>, large earthquakes would be predicted to return only once every 1,000 years, or even less frequently. The way out of this conundrum is not clear. There is no evidence for particularly frequent episodes of large aseismic slip in that area, and postseismic afterslip, although significant<sup>1</sup>, is much too small to balance the slip budget. So either the slip deficit accumulating in the interseismic period is overestimated (which might happen

if, for example, a fraction of interseismic strain is not recoverable), or it is incorrect to assume that geodetic rates measured over a decade or so are representative of strain build-up over periods of centuries to millennia.

Another paradoxical and possibly related observation is that the Tohoku-Oki earthquake induced more than 1 metre of systematic coastal subsidence, whereas uplift would have been expected to balance the subsidence rate of 5 millimetres per year during the interseismic period. The long-term coastal uplift requires deformation events that are large and frequent enough to compensate for that subsidence. This might call for a review of both the assumption that interseismic deformation of the upper plate is purely elastic, and the corollary that interseismic elastic strain is relaxed only by earthquakes that occur along the plate interface.

Finally, the geodetic data acquired both before<sup>5,7,8</sup> and after<sup>1</sup> the Tohoku-Oki earthquake suggest that the plate interface south of the rupture area is mostly creeping aseismically. There is thus no indication of a major zone of strain build-up on that portion of the plate boundary that might threaten Tokyo. But it is clear that although geodetic networks are

invaluable instruments for observing strain accumulation and seismic release at plate boundaries and major faults, we don't yet have an adequate theory to use these data for earthquake and tsunami hazard assessment. ■

**Jean-Philippe Avouac** is in the Division of Geological and Planetary Sciences, California Institute of Technology, Pasadena, California 91125, USA.

e-mail: avouac@gps.caltech.edu

1. Ozawa, S. *et al.* *Nature* doi:10.1038/nature10227 (2011).
2. Sato, M. *et al.* *Science* doi:10.1126/science.1207401 (2011).
3. Simons, M. *et al.* *Science* doi:10.1126/science.1206731 (2011).
4. Moreno, M. S., Bolte, J., Klotz, J. & Melnick, D. *Geophys. Res. Lett.* **36**, L16310, doi:10.1029/2009gl039276 (2009).
5. Suwa, Y., Miura, S., Hasegawa, A., Sato, T. & Tachibana, K. *J. Geophys. Res.* **111**, B04402, doi:10.1029/2004JB003203 (2006).
6. Savage, J. C. *Annu. Rev. Earth Planet. Sci.* **11**, 11–43 (1983).
7. Hashimoto, C., Noda, A., Sagiya, T. & Matsu'ura, M. *Nature Geosci.* **2**, 141–144 (2009).
8. Loveless, J. P. & Meade, B. J. *J. Geophys. Res. Solid Earth* **115**, B02410, doi:10.1029/2008jb006248 (2010).
9. Ide, S., Baltay, A. & Beroza, G. C. *Science* doi:10.1126/science.1207020 (2011).
10. Sawai, Y. *et al.* *Holocene* **18**, 517–528 (2008).